

# A Comprehensive Survey of Recent Proposed Content Replacement Strategies for Cooperative Edge Caching in IoT

Fariba Majidi<sup>1,2</sup> 

1- Department of Computer Engineering, Mobarakeh Branch, Islamic Azad University, Isfahan, Iran.  
Email: F.majidi@khuisf.ac.ir (Corresponding author)

2- Department of Computer Engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran.

## ABSTRACT:

The Internet of Things significantly increases the number of terminals and network traffic load, while its real-time applications require minimal latency to access the requested contents from data centers. Despite the processing and storage capabilities of base stations in 5G networks, the use of edge caching has proven to be an effective solution to reduce content access delay and repetitive traffic. This optimization of content transfer through the internet is crucial for maintaining the efficiency and performance of IoT applications. However, several challenges must be addressed. The limited storage resources, the constant changes in network nodes, and the dynamic patterns of content requests and user behavior pose fundamental challenges to content placement strategies. Developing an effective content placement strategy requires a comprehensive understanding of these factors and the ability to adapt to the evolving network environment. In this paper, Challenges and issues facing content placement strategies in cooperate edge caching are described and recent researches in this field are reviewed, categorized, and explained comprehensively.

**KEYWORDS:** Internet of Things, Edge caching, Edge computing, Deep Reinforcement Learning, Federate learning

## 1. INTRODUCTION

In recent years, With the advent of the Internet of Things (IoT), phones are not the only mobile terminals transmitting data, but numerous varieties of sensors are connected too. Therefore, information is not exchanged only between humans, but also is extended to communication between objects and can fulfill needs such as identification, location, tracking, management, and intelligent monitoring. Therefore, IoT significantly increases the number of terminals and network traffic load. IoT systems have presented new needs for cloud computing-based solutions. Especially in some real-time applications, such as smart cars, real-time processing of sensed data and executing the reactions is indispensable. Although the delay of communications within the local network is low, accessing the content available in data centers through the Internet has posed an important challenge to these systems [2].

Edge computing has brought data storage, computing, and control closer, in the edge devices rather than in a central cloud server. Therefore, each edge device plays its role in determining what information should be stored or processed locally and what should be retrieved from the cloud server. Edge computing has perfected the IoT in high scalability, low latency, location awareness, and real-time use of local devices' computing capabilities. As presented in [33], most network traffic for IoT systems consists of frequent requests for duplicate content. The continuous transfer of these popular contents will cause a significant increase in network traffic redundancy [65]. Therefore, by storing them in the cache of edge devices, the contents would be provided to end users without repeated transmission from the central servers [3]. Fig.1 shows how Edge caching would be implemented in a hybrid network. However, edge caching needs

Paper type: Research paper

<https://doi.org/xxx>

Received: 22 July 2024; revised: 5 August 2024; accepted: 11 September 2024; published: 1 December 2024

How to cite this paper: F. Majidi, "A Comprehensive Survey of Recent Proposed Content Replacement Strategies for Cooperative Edge Caching in IoT", *Majlesi Journal of Telecommunication Devices*, Vol. 13, No. 4, pp. 191-197, 2024.

some considerations. One of the issues is managing the resource utilization. The memory and processing resources of edge devices are limited. So, each node can only store a small amount of media content. Another important issue is network reliability which determines whether a network remains functional when its elements fail at random [5].

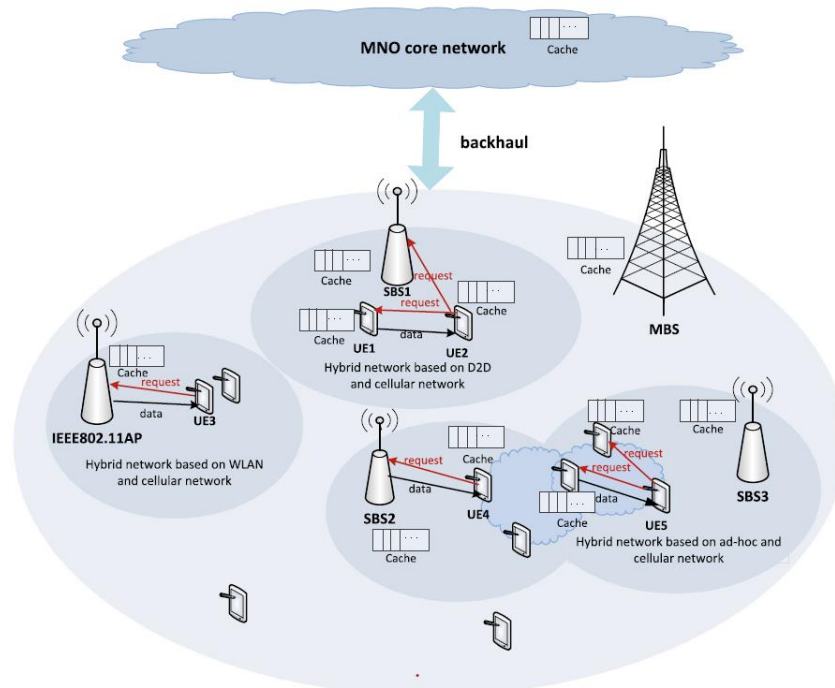


Fig. 1. Edge caching in a hybrid network [4].

Wide and fast communication of edge devices in the 5G network, especially interconnected BSs using the wired mesh network, has made caching possible cooperatively and intelligently. By cooperating between edge devices, it is possible to respond to content requests with local communications between BSs and users, significantly reducing content access delay, also it optimizes resource utilization and reliability.

In all approaches for increasing the performance of edge caching, content replacement strategy has been the focus of the recent edge cache research. If the popular contents cached in the edge devices are chosen accurately for feature requests, it will improve the hit rate of the edge cache. Due to the limitation of the cache space, a solution must be provided to select the contents for eviction when the cache is full. So that there would be enough space to insert new popular content. The replacement strategy is usually based on content popularity or the cost of storing and transferring them. In this paper, we survey recent state-of-the-art literature on cooperate edge caching and taxonomy based on the methods utilized for content replacement strategy as optimization methods and intelligent methods.

For the rest of this paper, we first describe the research that used optimization methods for solving the problem of content replacement and categorize them by the metrics they used for optimization formulas. Then the recently proposed intelligent methods as content replacement strategies are explained. The research that utilizes the most trends machine learning approach is classified in this section. Finally, we present the future outlook for this area and conclude the article.

## 2. OPTIMIZATION METHODS

In the optimization methods, to determine the appropriate content replacement strategy in the edge cache, the factors and problems to be optimized must be carefully selected. Also, the method of problem formulation has a great impact on the practical application and computational complexity of the proposed method because neglecting an important limitation can lead to solutions that are not applicable in practice [35]. The performance of edge cache is usually evaluated from two aspects: QoE and network performance. In most cases, hit rate and content access delay metrics are used for the QoE evaluations, also local network traffic load and request load on the central server, are considered to evaluate network efficiency. In the collaborative method, the classification factors Influence the proposed strategy considerably. In the rest of this section, content access delay, cache hit rate, resource utilization, cost, and classification factors are described as metrics mostly used in optimization methods.

## 2.1. Content access delay

One of the most important metrics to evaluate the efficiency of edge caching is the Content access delay [35]. Although this parameter depends on the processing time, but the network delay exerts a substantial impact on it. In the article [12], Saputra et al try to reduce the delay by formulating the contents that are stored in the cache and deciding how the devices access the content, then this problem has been converted into MINLP and it is solved with the divide and conquer algorithm. In [15], the problem is formulated as MILP, and in [20], an algorithm using belief propagation is presented to solve the proposed problem. In [37], the problem has been formulated as a multi-agent multi-armed bandit and has used replacement to solve this problem. In the article [38], the optimization problem of content storage in edge cache is formulated. So, the delay of content requests of all users in MEC networks could be minimized.

## 2.2. Edge cache hit rate

The cache hit rate is the probability that the content requested by the users exists in the cache of the edge devices [39]. In other words, the percentage of requests that will be fulfilled by the cache. Improving this parameter often has a positive effect on other parameters as well. Hence, optimization of cache hit rate has been considered as the main issue in many researches. In [14], the problem is formulated as MDP, and the solution is proposed by the placement strategy. In [17], Lagrangian multiplication and zero and one knapsack methods have been used to solve the problem, and in [7], the optimization of the hit rate of the edge cache in the Vehicular content network has been investigated, and in [40], the hit rate is optimized in three layers (routers, base stations, and users) also as interlayer and extra layer.

## 2.3. Cost

By reducing energy consumption, increasing power, and improving QoE, the costs of service providers are reduced and they can earn more with the same workload [39]. Answering popular requests locally also allows higher layers to handle more requests, improving overall performance. So, [44] solved the problem using graph coloring. In [41], one of the important optimization factors is energy, which is formulated as ILP. In [45], the optimization of the transmission cost is achieved using the PSO algorithm.

## 2.4. Device classification factors

As it was mentioned before, the classification BSs has a great impact on the efficiency of the collaborative edge cache, therefore [1], optimize the caching performance by finding the optimal distance between the base stations of a group and reducing the cooperation costs between them, and [11] has optimized the size of groups of cooperating BSs in their proposed merging optimization method.

It should be noted that due to the high complexity of edge deployment and the fully dynamic conditions of wireless networks, it is difficult to find a design-based solution [22]. However, even though some of the proposed methods that used optimization to improve QoE, have achieved good results, these methods in most cases have overlooked the long-term impact of current decisions. Therefore, a considerable number of the presented algorithms operate optimally or close to optimally in the system only for a specific period of time. Therefore, these methods cannot fully and effectively improve the efficiency of the edge cache [9]. A secondary section heading is enumerated by a capital letter followed by a period and is flush left above the section. The first letter of each important word is capitalized and the heading is italicized.

## 3. INTELLIGENT METHODS

Designing a suitable strategy for replacing contents in the collaborative edge cache requires considering the features and complexities of edge networks and the continuous mobility of wireless devices. If the proposed method relies on specific information as input, it may be difficult to obtain this information due to the great diversity in border channels and devices as well as security policies.

Another issue that should be considered when designing an algorithm for content replacement is the dynamics of content popularity. In each set of cooperating edge devices, the popularity of each content changes continuously, and content may be popular only in a short period of time [1]. Therefore, the placement strategy is optimal when it is based on the correct distribution of content popularity, by which it can recognize the content that will be requested in the near future. But predicting content popularity is difficult for several reasons:

The contents are requested by different users, so the popularity of the contents changes with the mobility of the local network users covered by each base station. Another problem is that users' interests vary in different situations (location, network topology, personality traits, etc.).

Recently, edge devices have higher computing and storage capabilities [9]. Therefore, it is possible to implement some methods such as massive data analysis and deep learning on these devices, and it is also possible to apply artificial intelligence techniques in mobile edge networks to understand user behaviors and network characteristics. With an

understanding of the user and the network, patterns can be designed that are context-aware and intelligent, allowing edge devices to make the right decision at the right time to choose what content to store in their limited cache resources [9].

Many recent researches have used strategies based on various artificial intelligence and machine learning methods for edge cache management. [4], [17], [24], [26]– [28], [46] In the next sections, methods using reinforcement learning, deep reinforcement learning, and federated learning for predicting user requests are expressed.

### 3.1. Reinforcement learning

Reinforcement learning is concerned with how an agent in an environment should act to maximize overall reward. This method is not a subset of any of the supervised and unsupervised methods [40]. The environment is usually modeled using Markov Decision Process (MDP), but instead of a precise mathematical model of MDP, the target of MDP is usually very large, and precise mathematical methods cannot be used for its design. فلانی و همکاران [47], using a history of content requests and multi-agent reinforcement learning, presented a method for replacing content in a collaborative edge cache.

### 3.1. Deep reinforcement learning

The use of reinforcement learning has allowed agents to solve the decision problem by learning from interacting with the environment. To achieve this goal, the environment must be determined in a suitable way and with an acceptable complexity, which makes the use of this method limited only to cases where the characteristics of the environment can be extracted. With the advent of deep neural networks, agents can learn some compact representations using high-dimensional and raw data. Deep learning is a type of artificial neural network that mimics the way the human brain works in data processing and pattern creation. Some recent research has used deep neural networks to model requested content. [6], [20], [28], [48] By combining deep learning and reinforcement learning, a powerful model can be created that has the ability to solve the great parts of previously intractable problems. Therefore, agents can have optimal control over the environment using the knowledge they obtain directly from the raw data. Researchers in [6], [13], [18], [19], [34], [28]–[31], [49], [50] utilize distributed and multi-agent deep reinforcement learning and have been able to significantly improve the hit rate by improved predicting the content in the collaborative edge cache.

### 3.2. Federated deep reinforcement learning

Another technique that has been proposed with the aim of improving security and reducing data transfer overhead in edge cache is federated learning. Federated learning is a method that uses agents that have local data samples to perform learning in a decentralized manner without transferring data to a central server or other agents. DRL techniques require a high computational capacity of resources to find optimal solutions. In particular, if there are a large number of resource optimization factors, parameters, and criteria for resource optimization in large-scale MEC systems (operator networks in cities), advanced distributed deep learning (DL) methods should be used for an operable method in the real world.

As shown in Fig.2, although maintaining and training each edge device as a DRL agent could increase the performance, the use of distributed DRL is only practical in MEC systems because of insufficient time and data for large-scale training. Also, most distributed DRL architectures cannot handle unbalanced and Non-IID data, and they can't support privacy issues too [32].

Therefore, the researchers in [8], [9], [22], [51] have used federated learning to train DRL agents. Because federated learning can overcome the following challenges as described in the following [32].

- Non-uniform distribution of data (Non-IID): The training data in the edge devices is based on the environment it has experienced. Each edge device has its own processing capability and energy consumption. Therefore, the local data of each of them cannot be a suitable representative for the training data of all end users and edge devices. In federated learning, this challenge can be overcome by integrating updated models with FedAvg. [32]
- Limited communications: Users often go offline unpredictably or have poor communication resources. Federated learning requires only a fraction of users to upload their updates during a training session, so it can handle situations where clients often go offline unpredictably [9].
- Imbalance: Some edge devices may have more computational tasks and some may experience more states of mobile networks, resulting in different amounts of training data among them. This challenge can be dealt with by the FedAvg algorithm [9].
- Privacy and Security: The amount of information that needs to be loaded for federated learning is the minimum volume of updates that can be used to improve the behavior of the DRL agent. In addition, different privacy-preserving and secure aggregation techniques can be applied, which prevent the inclusion of privacy-sensitive information in local updates [32].

Deep Learning with Edge		
<b>DRL</b>	<b>Distributed DRL</b>	<b>Federated Learning</b>
Pros: Best Performance	Pros: Fast Training; Barely Usable in Edge	Pros: Minimum Data Transmission; Privacy Protection; Flexible Training; Robust to Unbalanced and non-IID data; ...
Cons: Impractical in Edge (Massive Redundant Data Transmission; Privacy Risk;)	Cons: Privacy Risk; Weaker Performance; ...	Cons: Near Best Performance
Small-Scale <ul style="list-style-type: none"> <li>• Caching Policy</li> <li>• Traffic Engineering ...</li> </ul> On One UE or Edge Node	Medium <ul style="list-style-type: none"> <li>• Resource Allocation</li> <li>• Caching Policy</li> <li>• Computation Offloading Policy</li> <li>• Traffic Engineering ...</li> </ul>	Large-Scale <ul style="list-style-type: none"> <li>• Resource Allocation</li> <li>• Caching Policy</li> <li>• Computation Offloading Policy</li> <li>• Traffic Engineering ...</li> </ul>

Fig. 2. Comparison of centralized, distributed, and federated methods in using DRL for border cache management [9].

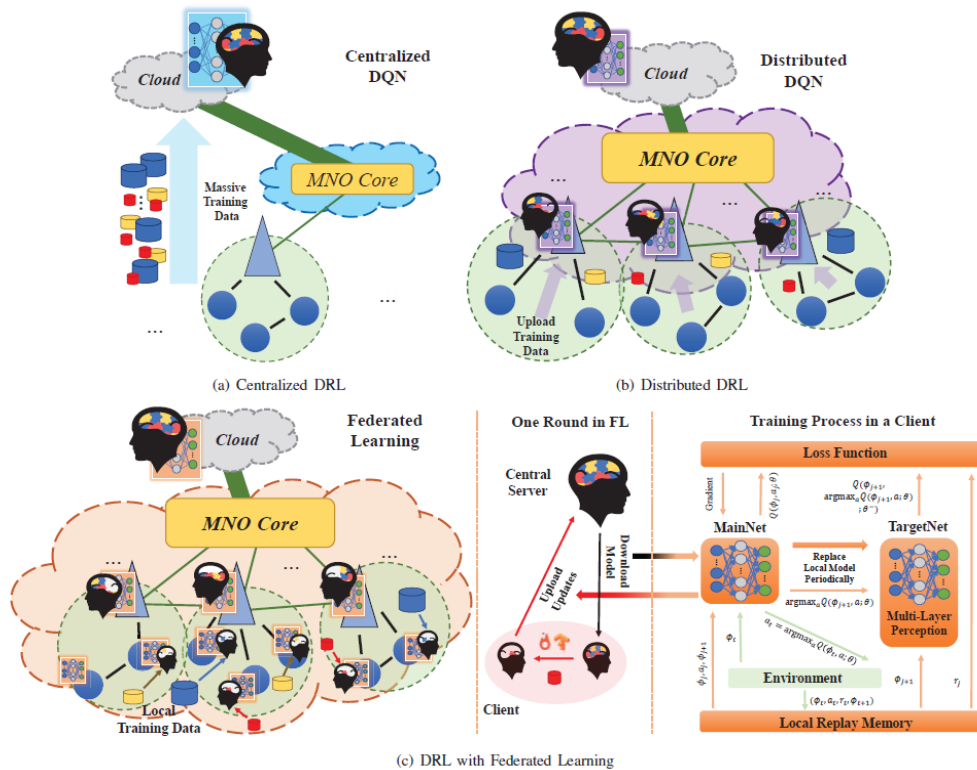


Fig. 3. Modeling of centralized, distributed and federated DRL methods presented in [9].

[8]-[9], [22] have proposed a method by combines federated learning and deep reinforcement learning. In these methods, each base station implements deep reinforcement learning using its local data and sends only the updated model to the central server. By aggregating the models received from a set of agents, the central server creates an updated model and returns it to all active agents to continue learning and predict popular content that they should store in the



collaborative edge cache. Figure 3-3 compares the method presented in [9] with centralized and distributed DRL methods. Fig.3 compares the method presented in [9] with centralized and distributed DRL methods.

#### 4. CONCLUSIONS

Recent research reviewed in this paper has shown two main challenges for improving cooperate edge cache performance. Correctly predicting future user requests with the lowest processing cost is the first and most important challenge. To overcome this challenge, despite the existing capacities of the current edge networks, the best solution is to use intelligent methods to determine the most suitable content replacement strategy in the edge cache. The second important challenge is the effective cooperation of edge devices in storing and responding to user requests. The numerous edge devices could utilize their storage for more variation of contents so respond locally to a relatively large number of user requests. However, their capability to cooperate in caching popular contents would increase the edge cache hit rate and pose a significant challenge to finding an optimal method for Determining cooperating groups and their cooperation methods so that, in addition to improving the hit rate of the cache, it does not impose too much traffic overhead on the local network, which slows down the user's access to content and causes longer delays. The problem of providing privacy and network security is also a challenge that should always be considered in multiple-device cooperation in a public network.

#### REFERENCES

- [1] L. Yang, Y. Chen, L. Li, and H. Jiang, *Cooperative Caching and Delivery Algorithm Based on Content Access Patterns at Network Edge*, vol. 278. Springer International Publishing, 2019.
- [2] Z. Piao, M. Peng, Y. Liu, and M. Daneshmand, "Recent Advances of Edge Cache in Radio Access Networks for Internet of Things: Techniques, Performances, and Challenges," *IEEE Internet Things J.*, vol. PP, no. c, p. 1, 2018.
- [3] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. C. M. Leung, "Cache in the air: Exploiting content caching and delivery techniques for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 131–139, 2014.
- [4] G. Xu et al., "An Incremental Learning Based Edge Caching System : From Modeling to Evaluation," vol. 8, pp. 12499–12509, 2020.
- [5] F. Majidi, M. R. Khayyambashi and B. Barekatin, "HFDRL: An Intelligent Dynamic Cooperate Caching Method Based on Hierarchical Federated Deep Reinforcement Learning in Edge-Enabled IoT," in *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 1402-1413, 15 Jan.15, 2022, doi: 10.1109/JIOT.2021.3086623.
- [6] R. Wang, M. Li, L. Peng, Y. Hu, M. M. Hassan, and A. Alelaiwi, "Cognitive multi-agent empowering mobile edge computing for resource caching and collaboration," *Futur. Gener. Comput. Syst.*, vol. 102, pp. 66–74, 2020.
- [7] Z. Su, Y. Hui, Q. Xu, T. Yang, J. Liu, and Y. Jia, "An edge caching scheme to distribute content in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 5346–5356, 2018.
- [8] Z. Yu et al., "Federated Learning Based Proactive Content Caching in Edge Computing," 2018 IEEE Glob. Commun. Conf. GLOBECOM 2018 - Proc., 2018.
- [9] X. Wang, Y. Han, C. Wang, Q. Zhao, X. Chen, and M. Chen, "In-Edge AI: Intelligentizing Mobile Edge Computing, Caching and Communication by Federated Learning," *IEEE Netw.*, vol. XX, pp. 1–10, 2019.
- [10] Z. Zheng, L. Song, Z. Han, G. Y. Li, and H. V. Poor, "A stackelberg game approach to proactive caching in large-scale mobile edge networks," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 8, pp. 5198–5211, 2018.
- [11] S. Zhang, P. He, K. Suto, P. Yang, L. Zhao, and X. Shen, "Cooperative Edge Caching in User-Centric Clustered Mobile Networks," *IEEE Trans. Mob. Comput.*, vol. 17, no. 8, pp. 1791–1805, 2018.
- [12] Y. M. Saputra, H. T. Dinh, D. Nguyen, and E. Dutkiewicz, "A Novel Mobile Edge Network Architecture with Joint Caching-Delivering and Horizontal Cooperation," *IEEE Trans. Mob. Comput.*, pp. 1–1, 2019.
- [13] Y. M. Saputra, D. T. Hoang, D. N. Nguyen, E. Dutkiewicz, D. Niyato, and D. I. Kim, "Distributed Deep Learning at the Edge: A Novel Proactive and Cooperative Caching Framework for Mobile Edge Networks,"
- [14] C. Wang, S. Wang, D. Li, X. Wang, X. Li, and V. C. M. Leung, "Q-learning based edge caching optimization for D2D enabled hierarchical wireless networks," *Proc. - 15th IEEE Int. Conf. Mob. Ad Hoc Sens. Syst. MASS 2018*, pp. 55–63, 2018.
- [15] N. Wang, G. Shen, S. K. Bose, and W. Shao, "Zone-Based Cooperative Content Caching and Delivery for Radio Access Network with Mobile Edge Computing," *IEEE Access*, vol. 7, pp. 4031–4044, 2019.
- [16] X. Li, X. Wang, P. J. Wan, Z. Han, and V. C. M. Leung, "Hierarchical edge caching in device-to-device aided mobile networks: Modeling, optimization, and design," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 8, pp. 1768–1785, 2018.
- [17] X. Zhao, P. Yuan, H. Li, and S. Tang, "Collaborative edge caching in context-aware device-to-device networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9583–9596, 2018.
- [18] X. He, K. Wang, and W. Xu, "in Edge-Enabled IoT," *IEEE Comput. Intell. Mag.*, vol. 14, no. NOVEMBER, pp. 10–20, 2019.
- [19] G. Qiao, S. Leng, S. Maharjan, Y. Zhang, and S. Member, "Deep Reinforcement Learning for Cooperative Content Caching in Vehicular Edge Computing and Networks," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 247–257, 2020.
- [20] Y. Zhang, Y. Li, R. Wang, J. Lu, X. Ma, and M. Qiu, "PSAC: Proactive Sequence-aware Content Caching via Deep Learning at the Network Edge," *IEEE Trans. Netw. Sci. Eng.*, vol. 4697, no. c, pp. 1–1, 2020.
- [21] W. Chien, H. Weng, and C. Lai, "Q-learning based collaborative cache allocation in mobile edge computing," *Futur. Gener. Comput. Syst.*, vol. 102, pp. 603–610, 2020.

- [22] X. Wang, C. Wang, X. Li, V. C. M. Leung, and T. Taleb, "Federated Deep Reinforcement Learning for Internet of Things with Decentralized Cooperative Edge Caching," *IEEE Internet Things J.*, no. c, pp. 1–1, 2020.
- [23] P. Wu, J. Li, L. Shi, M. Ding, K. Cai, and F. Yang, "Dynamic Content Update for Wireless Edge Caching via Deep Reinforcement Learning."
- [24] A. Mehrabi, M. Siekkinen, and A. Yla-Jaaski, "QoE-Traffic Optimization Through Collaborative Edge Caching in Adaptive Mobile Video Streaming," *IEEE Access*, vol. 6, pp. 52261–52276, 2018.
- [25] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, 2014.
- [26] M. F. Pervej, L. T. Tan, and R. Q. Hu, "Artificial Intelligence Assisted Collaborative Edge Caching in Small Cell Networks," *arXiv*, 2020.
- [27] N. Garg et al., "Online Content Popularity Prediction and Learning in Wireless Edge Caching," vol. 68, no. 2, pp. 1087–1100, 2020.
- [28] H. Pang, J. Liu, X. Fan, and L. Sun, "Toward Smart and Cooperative Edge Caching for 5G Networks: A Deep Learning Based Approach," 2018 IEEE/ACM 26th Int. Symp. Qual. Serv. IWQoS 2018, 2019.
- [29] J. Zheng et al., "Smart Edge Caching-Aided Partial Opportunistic Interference Alignment in HetNets," *Mob. Networks Appl.*, vol. 25, no. 5, pp. 1842–1850, 2020.
- [30] M. C. Gursoy, C. Zhong, and S. Velipasalar, "Deep Multi-Agent Reinforcement Learning for Cooperative Edge Caching," pp. 439–457, 2020.
- [31] F. Wang, F. Wang, J. Liu, R. Shea, and L. Sun, "Intelligent Video Caching at Network Edge : A Multi-Agent Deep Reinforcement Learning Approach," 2017.
- [32] H. Brendan McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, "Communication-efficient learning of deep networks from decentralized data," *Proc. 20th Int. Conf. Artif. Intell. Stat. AISTATS 2017*, 2017.
- [33] X. Sun and N. Ansari, "Dynamic Resource Caching in the IoT Application Layer for Smart Cities," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 606–613, 2018.
- [34] Z. Piao, M. Peng, Y. Liu, and M. Daneshmand, "Recent Advances of Edge Cache in Radio Access Networks for Internet of Things: Techniques, Performances, and Challenges," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 1010–1028, 2019.
- [35] L. Li, G. Zhao, and R. S. Blum, "A survey of caching techniques in cellular networks: Research issues and challenges in content placement and delivery strategies," *IEEE Commun. Surv. Tutorials*, vol. 20, no. 3, pp. 1710–1732, 2018.
- [36] J. Liu, B. Bai, J. Zhang, and K. B. Letaief, "Content caching at the wireless network edge: A distributed algorithm via belief propagation," 2016 IEEE Int. Conf. Commun. ICC 2016, 2016.
- [37] W. Jiang, G. Feng, S. Qin, and Y. Liu, "Multi-Agent Reinforcement Learning Based Cooperative Content Caching for Mobile Edge Networks," *IEEE Access*, vol. 7, pp. 61856–61867, 2019.
- [38] Z. Sang, S. Guo, Q. Wang, and Y. Wang, "GCS: Collaborative video cache management strategy in multi-access edge computing," *Ad Hoc Networks*, vol. 117, 2021.
- [39] S. Dutta and A. Narang, "Dynamic Uncertainty-Based Analytics for Caching Performance Improvements in Mobile Broadband Wireless Networks," *Big Data Princ. Paradig.*, pp. 389–415, 2016.
- [40] X. Zhang and Q. Zhu, "Collaborative hierarchical caching over 5G edge computing mobile wireless networks," *IEEE Int. Conf. Commun.*, vol. 2018-May, pp. 1–6, 2018.
- [41] D. Ren, X. Gui, K. Zhang, and J. Wu, "Hybrid collaborative caching in mobile edge networks: An analytical approach," *Comput. Networks*, vol. 158, pp. 1–16, 2019.
- [42] T. X. Tran, P. Pandey, A. Hajisami, and D. Pompili, "Collaborative multi-bitrate video caching and processing in Mobile-Edge Computing networks," 2017 13th Annu. Conf. Wirel. On-Demand Netw. Syst. Serv. WONS 2017 - Proc., pp. 165–172, 2017.
- [43] Z. Chen, J. Lee, T. Q. S. Quek, and M. Kountouris, "Cooperative Caching and Transmission Design in Cluster-Centric Small Cell Networks," *IEEE Trans. Wirel. Commun.*, vol. 16, no. 5, pp. 3401–3415, 2017.
- [44] L. Chen and J. Xu, "Collaborative Service Caching for Edge Computing in Dense Small Cell Networks," pp. 1–30, 2017.
- [45] S. Wang, X. Zhang, K. Yang, L. Wang, and W. Wang, "Distributed Edge Caching Scheme Considering the Tradeoff Between the Diversity and Redundancy of Cached Content," 2017.
- [46] M. K. Somesula, R. R. Rout, and D. V. L. N. Somayajulu, "Contact duration-aware cooperative cache placement using genetic algorithm for mobile edge networks," *Comput. Networks*, vol. 193, 2021.
- [47] A. Sadeghi, F. Sheikholeslami, and G. B. Giannakis, "Optimal and Scalable Caching for 5G Using Reinforcement Learning of Space-Time Popularities," *IEEE J. Sel. Top. Signal Process.*, vol. 12, no. 1, pp. 180–190, 2018.
- [48] Y. Dai, D. Xu, S. Maharjan, G. Qiao, and Y. Zhang, "Artificial Intelligence Empowered Edge Computing and Caching for Internet of Vehicles," *IEEE Wirel. Commun.*, vol. 26, no. 3, pp. 12–18, 2019.
- [49] W. Shi et al., "LEAP: Learning-based smart edge with caching and prefetching for adaptive video streaming," *Proc. Int. Symp. Qual. Serv. IWQoS 2019*, 2019.
- [50] I. A. Elgendy, W. Z. Zhang, H. He, B. B. Gupta, and A. A. Abd El-Latif, "Joint computation offloading and task caching for multi-user and multi-task MEC systems: reinforcement learning-based algorithms," *Wirel. Networks*, vol. 27, no. 3, pp. 2023–2038, 2021.
- K. Qi and C. Yang, "Popularity Prediction with Federated Learning for Proactive Caching at Wireless Edge," *IEEE Wirel. Commun. Netw. Conf. WCNC*, vol. 2020-May, 2020.

# Power Divider Design in the Silicon Photonic Integrated Substrate (SPIC) Technology with Purpose of the Signal Transmission Coefficient in the 5G Band Frequency

Maryam Bayati<sup>1</sup> , Alireza Ghasemi<sup>2</sup>

1- Sepahan Institute of Higher Education, Science & Technology, Faculty of Technical engineering, Isfahan, Iran.  
Email: m.bayati@sepahan.ac.ir (Corresponding author)

2- Department of Electrical Engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran.  
Email: alireza.ghasemi.isf@gmail.com

## ABSTRACT:

In research, the signal Power Divider (PD) has been designed with a photonic integrated substrate with the silicon substrate and benefits from MEMS technology in its geometry configuration in the microwave high-frequency bandwidth of 120 to 260 GHz with coverage of signal exchange with telecommunication bulk data in the fifth-generation protocol been done. The proposed has been designed micrometer dimensions with wide microwave bandwidth coverage in the division of bulk data signals, which indicates the use of this model in optical telecommunication devices. Also, it has been used by BCPD, TL, WPD, and PNPTL techniques to create a configuration with a simple geometry in the form of a U-shape and a T-shape connection, each of the branches is isolated in very small dimensions, and impedance matching and creates frequency balance in signal exchange and division; So it has been results obtained from it includes the improvement of the dispersion parameter and also causes its time delay function to have a lower value of the time delay function than the time derivability function to the frequency bandwidth in the sigma function in the differential order, which has been being around 4.3 (ns) in the range of 170GHz has been obtained in the value of -3.1 (dB), and the value of its transmission coefficient is also in the range of -3.0 to -4.98 (dB). The design has been our technique for improving the numerical values of the dispersion parameters and creating a configuration in the simplicity of its geometry in dimensions  $T=10\mu\text{m}$ ,  $L=315\mu\text{m}$ , and  $W=400\mu\text{m}$ . The application of the proposed has been used in creating a photonic substrate in signal transmission for high bandwidth microwave in the 5G telecommunication protocol, which is simulated in the ADS software.

**KEYWORDS:** Power Divider, MEMS, PIC, 5G-Microwave, Silicon.

## 5. INTRODUCTION

With the design of new circuits and the use of signal dividers and collectors in telecommunication circuits, the increasing need of telecommunication systems for signal dividers and collectors is fulfilled. Due to this, by creating a signal divider and collector platform in the receiver and transmitter circuits, we can perform signal divider and collector in different circuits and to be able to send signals with a suitable platform for dividing and collecting signals so that the amount of delay and noise is low. By creating a suitable substrate, the noise of the substrate can be minimized. It is also possible to design the signal divider and collector as a circuit element with admittance and impedance structure; But in the majority of signal divider and collector circuits are designed in passive mode and transistor signal switching in active mode, which are designed in different microwave frequency bands.

## 6. TECHNICAL WORK PREPARATION

The proposed design is in the design of the passive signal splitter and collector model in the frequency bandwidth of 5 GHz high band microwave protocols which is simulated in the material under the silicon layer by considering pic and RF MEMS technologies by having several signal ports in the ads software. One of the users of this research is the application between the antenna and the telecommunication transmitter board, which is used in portable and stationary



devices. Another use of it is to use it in the communication between the internal structure of signal amplifiers hampon circuits, which are designed in several layers, which are designed in the input part of the amplifier circuit from the signal divider bed and at its output from the signal collector bed. In the transmission of the signal from an input port to several separate branches to divide the signal power equally, and in collecting the input signals, separate branch paths are used to the output port. Time delay function or signal switching circuit is used to determine the signal division and addition function. So that the substrate with two functions is placed between the antenna and the circuit or between two blocks of the circuit and by determining the signal exchange in the specified time function, the frequency signal transmission is done.

### **6.1. Designing a signal splitter substrate with silicon substrate material in the field of photonics**

The research conducted is about electrophotonics, which is carried out in the field of microwave circuits. Many applications of photonic devices and technologies for circuits of arbitrary microwave waveforms, with high processing speeds, use microwave circuits in the field of optoelectronics. This class of photonic applications for microwave engineering is known as microwave photonics. Extensive capabilities have enabled the realization of key functions that are either very complex or simply not possible in the microwave domain alone. Recently, this growing field has adopted integrated photonics technologies to develop microwave photonic systems with greater robustness as well as significant reductions in size, cost, weight, and energy consumption. In particular, silicon photonics technology is of great interest for this purpose because it offers tremendous possibilities for the integration of highly complex active and passive photonic devices, allowing seamless integration of MWP with high-speed silicon electronics [1].

### **6.2. The design of the signal splitter substrate with the aim of improving the return loss in the WIFI, LTE frequency band**

In this research, we present the first experimental demonstration of a microwave photonic integrated circuit for analog active interference cancellation. This circuit is unique in its ability to operate in any radio frequency band from 400 MHz to 6 GHz while requiring no optical input or output. First, we examine the amount of interference cancellation that can be achieved over a wide range of operating frequencies. We show that the circuit can achieve close to -30 dB of interference cancellation in all available frequency division duplex local thermal balance bands and WIFI bands. Second, we examine aspects of integrated circuit control and determine how much amplitude and phase adjustability can be built into the implementation [2].

### **6.3. Designing a signal transmission substrate with the purpose of a low-pass filter in optical circuits**

In this research, silicon-based photonic integration, there is a growing interest in the electronic circuit community to develop hybrid photonic electronic systems. However, photonic integrated circuit design tools using numerical methods are focused on solving Maxwell's equations. PIC system-level design tools recently use s-parameter modeling of optical components. However, detailed modeling of electronic driver and interface circuits is not supported on such platforms [3].

### **6.4. Design of wideband signal splitter on silicon substrate**

A new method for near-hermetic packaging of integrated passive UWB devices based on multilayer liquid crystal polymer technology is presented in this research. The aim of this work is to develop an inexpensive and easy-to-fabricate method that can be used to encapsulate already designed devices without the need for redesign and operate up to 20 GHz. To achieve this goal, a coplanar waveguide-to-microstrip transmission is designed to connect the enclosed device with the external environment. This allows the device to be mounted to the PCB and also isolated [4].

### **6.5. Signal division substrate in creating integration with CMOS substrates**

In this research, we are dealing with the design of the integration platform of photonics and electronics in the transceiver, the energy efficiency of broadband acceleration and a path for radical miniaturization has been facilitated. We present and implement a wafer-to-wafer integration method that combines electronic and photonic casting technologies. The motivation to integrate electronic and photonic systems into an aligned manufacturing process is long-standing, but it is now becoming a necessity with light-based detection and ranging subsystems envisioned for future data center technologies And high-density programmable neuromorphic photonics in the future of high-speed, high-density, low-energy electrical interconnects in transceivers also enable performance improvements through optical serialization advance equalization and drive voltage reduction. Photonic transmitters require accurate dc biasing and broadband impedance matching to ensure data integrity. Photonics and electronics chipsets, historically designed separately, have been implemented with biased tees and 50-ohm interfaces to facilitate interoperability and portable

design, but with increasing speed in a limited energy envelope this is no longer attractive. Components must be placed closer together within the package towards a composite 3D assembly and ultimately on the same processed wafer.

## 7. APPENDIX

The proposed design is used in the integrated platform with the ability to use the passive signal splitter model in photonics and telecommunication circuits, which by using MEMS technology has caused the geometry of the proposed design to be designed in micrometer dimensions. The use of mems technology in the design of signal exchange platforms makes its configuration simple and its dimensions are very small. This causes the proposed model to be designed in a very small geometry and the efficiency of the circuit increases. In substrate models, the feature of integration is used to reduce signal transmission attenuation at the substrate level for dividing and exchanging telecommunication signals. By using OIC, MEMS technologies, this causes signal transmission to have the least delay and attenuation, and the results of its dispersion parameters are also improved. This is one of the important features in the design at high microwave frequencies. **Table 1** shows the values of the standard parameters in measuring the results of various signal splitter design technologies. Also, in **Table 2**, the specifications of the frequency bandwidth in the 5g protocol are shown.

**Table 1.** Values of standard parameters in measuring the results of various signal splitter design technologies.

Parameter	Value	Value of Range	Frequency Coverage	Device Passive/Active
S11 (dB)	Under (-0)	[-0 ~ -30]	All microwave frequency bands	ALL
S12 (dB)	Under (-30)	[-0 ~ -40]	All microwave frequency bands	ALL
S21 (dB)	-20 to 0	[-0 ~ -5]	All microwave frequency bands	ALL
S22 (dB)	Under (-0)	[-0 ~ -30]	All microwave frequency bands	ALL
Delay Time (ns)	[-,+]	1nS, [-60 ~ +60]	All microwave frequency bands	ALL
VSWR (s/v)	UP +0	[+0 ~ +80]	All microwave frequency bands	ALL
Isolation (dB)	Under (-5)	[-0 ~ -35]	All microwave frequency bands	ALL

**Table 2.** Specification of frequency bandwidth in 5g protocol.

Protocol	Frequency Channels	Frequency Bands	Application
5G, multi-bands	5	5	Device, Radars
<b>STD, Down-Band. 1</b>	<b>STD, Down-Band. 2</b>	<b>STD, Down-Band. 3</b>	<b>IEEE Legal</b>
3.0-50.0 GHz	50.0-100.0 GHz	100.0-250.0 GHz	RF, microwave, millimeter-wave

## 8. RESULTS AND DISCUSSION

In this article, the analysis and dissemination of the design process of the proposed divider has been discussed in order to improve the parametric quantities measured in the simulation. Due to the fact that the splitter design model is capable of three ports to connect to the antenna and the telecommunication transmitter and receiver system and between circuits, in its technical design structure, the splitter is used in the high frequency bandwidth of the microwave in the high frequency volume data range of navigation. Because of this proposed design of the relevant devices in the side band, you can easily exchange signals in telecommunication radars and 5g frequency bandwidth. This improves the coefficient of dispersion parameters and related quantities of this parameter in radars set in the frequency band adjacent to the standard band. This research has been devoted to presenting a new model of the divider structure with PIC technology with the aim of application in the industry of telecommunication radars and optical devices. The design of the mems divider is done with the purpose of exchanging the signal transmission in the high frequency band. The geometry of the proposed design is U-shaped, and its connection is made in T-shape, which is spread in the dimensions of micrometers in the propagation of electromagnetic waves in E-plane and H-plane on the substrate surface of its silicon rough layer and it is shown in figures 1 to 7. From the proposed design with the aim of reducing the dimensions of the model and improving the coefficient of dispersion parameters and reducing the time delay coefficient for the applications

of microwave high frequency band radars in the exchange of bulk data over long distances in the 5g telecommunication protocol and optical telecommunication devices.

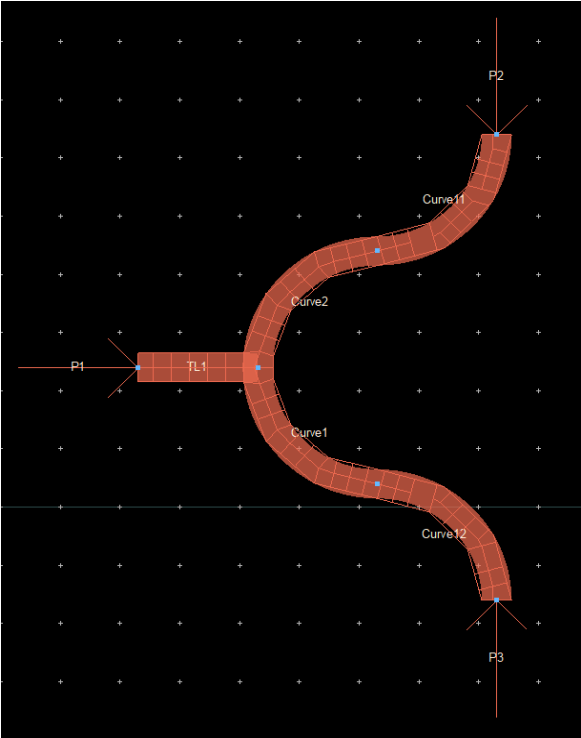


Fig. 1. Signal splitter schematic.

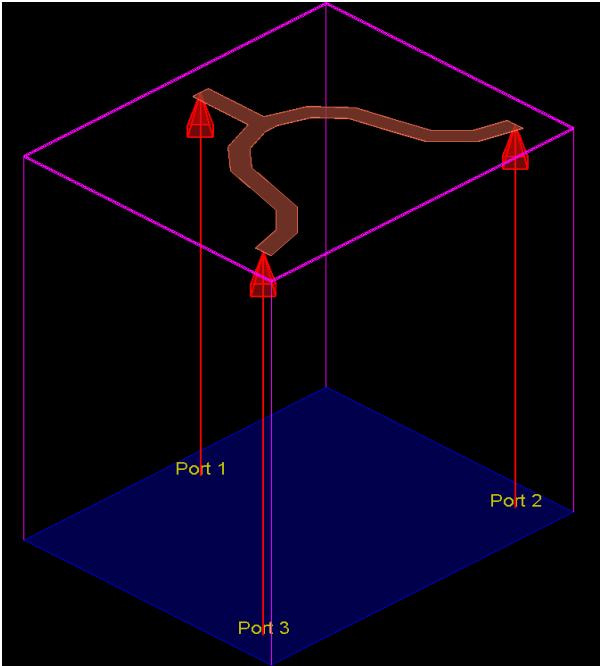
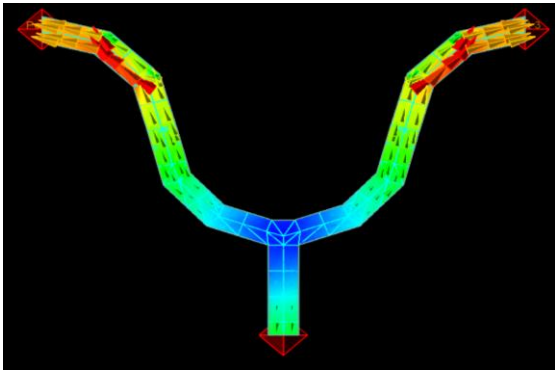
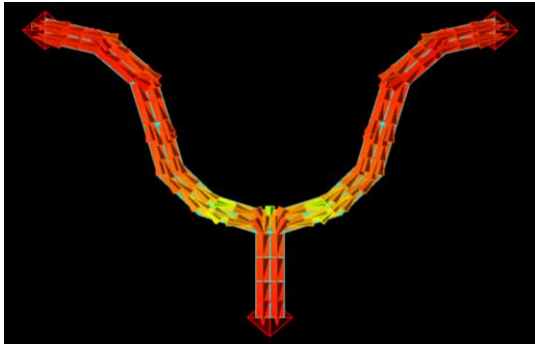


Fig. 2. 3D view of the proposed design

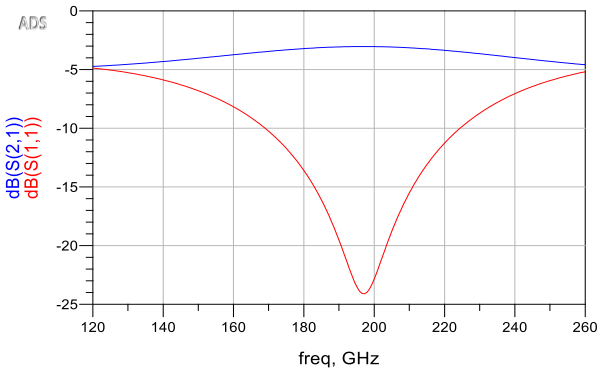


(a)

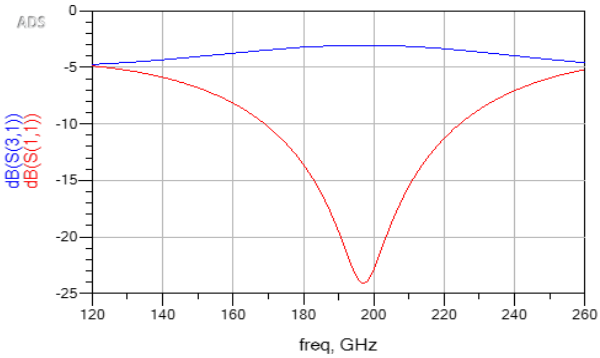


(b)

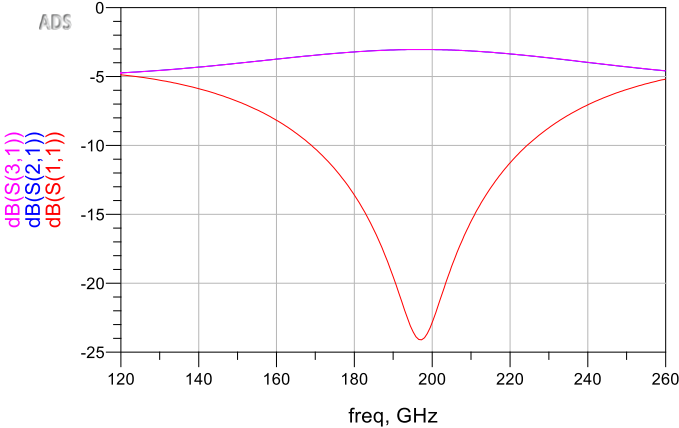
Fig. 3. (a) and (b) Signal transmission on the surface of the substrate in the form of convergence of electromagnetic waves.



(a)

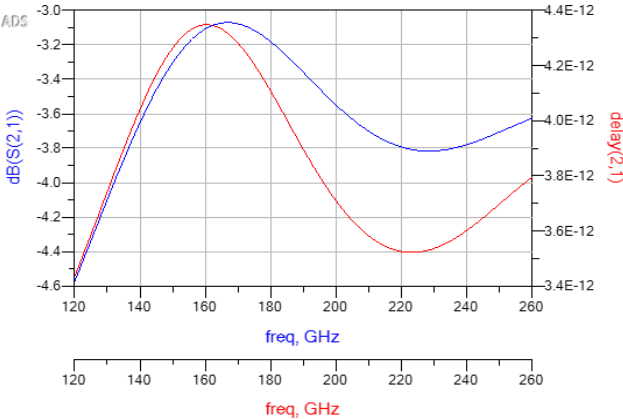


(b)

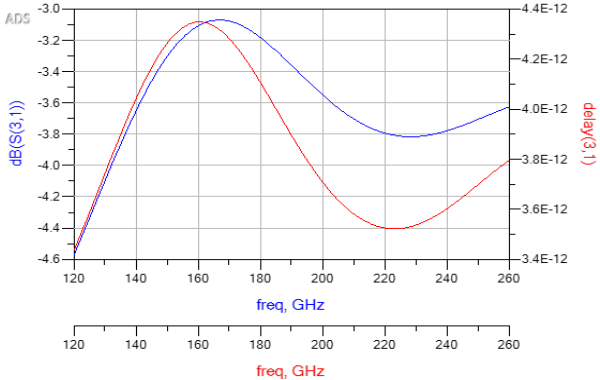


(c)

Fig. 4. Signal transmission coefficient diagram for each port.



(a)



(b)

Fig. 5. Time delay function diagram in signal transmission for each port.



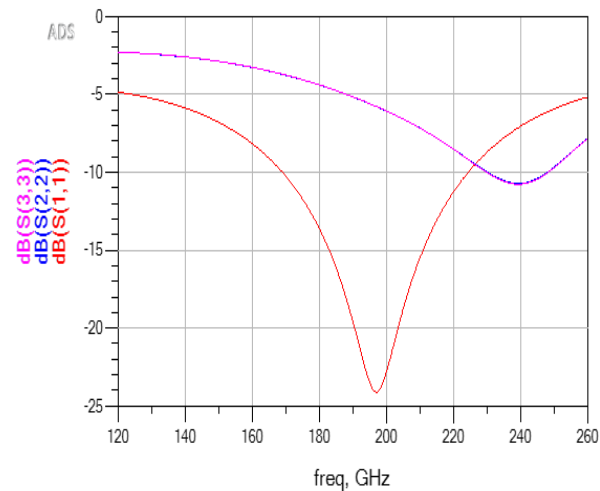


Fig. 6. Return loss diagram and output coefficient reflection.

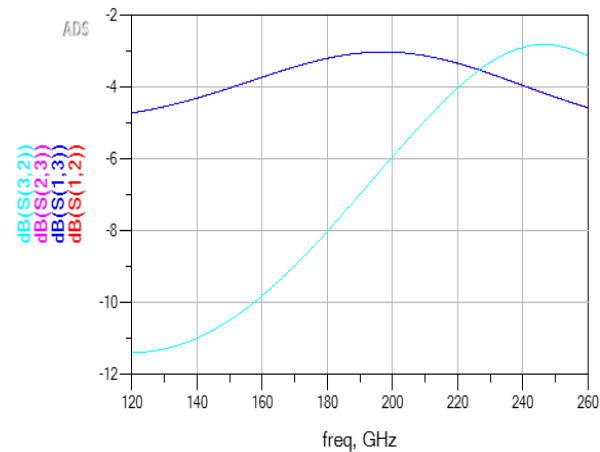


Fig. 7. Diagram of input to output isolation coefficient - output to output.


## 9. ACKNOWLEDGMENT

The research that has been done in the field of RF electronics is for the integrated and integrated design of the passive model of the signal power divider and the passive model of the signal collector. According to the base models, electronic resistors are designed in several ways in different technologies and for use in the system.

## REFERENCES

- [51] H. -Y. Liu, J. Xie and K. -K. M. Cheng, "Impact of Nonideal Auxiliary Current Profile on Linearity of Microwave Doherty Amplifiers: Theory and Experiments," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, DOI: 10.1109/TCSI.2022.3152589.
- [52] A. Jarndal and K. H. Hamza, "GaN Power Amplifiers Design Using Efficient GA-ANN Dynamic Nonlinear Model," 2021 14th International Conference on Developments in eSystems Engineering (DeSE), 2021, pp. 413-417, DOI: 10.1109/DeSE54285.2021.9719341.
- [53] G. Bosi, A. Raffo, R. Giofrè, V. Vadalà, G. Vannini and E. Limiti, "Empowering GaN-Si HEMT Nonlinear Modelling for Doherty Power Amplifier Design," 2020 15th European Microwave Integrated Circuits Conference (EuMIC), 2021, pp. 249-252, DOI: 10.1109/EuMIC48047.2021.00074.
- [54] M. Fares, D. Y. Wu, S. Boumaiza and J. Wood, "Inward nonlinear characterization of Doherty Power Amplifiers," 2009 IEEE MTT-S International Microwave Symposium Digest, 2009, pp. 1545-1548, DOI: 10.1109/MWSYM.2009.5166004.

# Proposing a Novel Method in Diagnosing Power Transformer Failures Based on the Analysis of Morphological Components

Amir ZamanVaziri<sup>1</sup>, Mehran Emadi<sup>2</sup>

1- Department of Electrical Engineering, Mobarkeh Branch, Islamic Azad University, Mobarkeh, Isfahan, Iran.  
Email: amir\_zamnvaziri@yahoo.com

2- Department of Electrical Engineering, Mobarkeh Branch, Islamic Azad University, Mobarkeh, Isfahan, Iran.  
Email: emadi.mehran49@gmail.com (Corresponding author)

## ABSTRACT:

Power transformers are very important in electrical energy distribution systems. In establishing a power plant and even in distribution networks, the cost of power transformers is significant. Any damage or failure in transformers can cause irreparable and heavy losses. Therefore, this equipment is always important in periodic visits. In periodic visits, various parameters of this equipment are checked. One of the ways used in periodic visits is to use infrared images to evaluate the health of this equipment. However, the analysis of these images by human experts is always error-prone and expensive. In this article, a new method is proposed in machine vision to diagnose transformer faults. In the proposed method, infrared images are pre-processed and then features based on multi-resolution transformations such as morphological component analysis, wavelet transform and bandlet transform are extracted and dimension reduced with the help of independent component analysis. The given one-dimensional features are classified using random forest classification, support vector machine and k-nearest neighbor. The classification accuracy obtained in the random forest bin class as the best classifier in fault detection is equal to 98.81%, as well as 91.51% sensitivity and 84.54% negative news rate, as well as 91.86% negative news rate compared to other the numbers showed their superiority. In the F criterion, this value has reached 0.99, which shows the efficiency of the proposed method.

**KEYWORDS:** Transformer, Failure, Analysis of Morphological Components, Image Processing.

## 10. INTORADUCTION

As the heart of a substation, the power transformer acts as an important link for voltage conversion and energy delivery. Survey results show that most of the power transformers around the world were installed in the 1980s and many of them are reaching the end of their life. The failure rate of these transformers has been continuously increasing in recent years, especially in power companies with poor maintenance and asset management methods, causing the loss of this valuable equipment [1]. Any change and failure in the transformer causes unwanted heat in this equipment. Hot spots in electrical equipment, especially power transformers, often occur as a result of looseness, oxidation and corrosion of connections, asymmetry of phases, and insulation failure of coils. All these things can be done by using thermal or infrared IR imaging equipment to measure the temperature of equipment such as energy distribution overhead lines, cables, transformers and fuses, cables, electrical panel equipment and all electrical distribution network equipment, including substations. Determined ground [2]. Diagnosing transformer failure is done online with the help of infrared image. These images are sensitive to the heat caused by various errors in the transformer. The performance of these cameras is limited by the level of image processing used. In other words, in infrared imaging in thermal cameras, the proposed method is important in detecting failures in image processing[3-6]. With the development and spread of substation inspection robots and smart substations in distribution networks, a large number of infrared fault images need to be analyzed. The infrared image transformer fault detection method can automatically identify

and analyze the collected infrared images, which can reduce the maintenance costs of transformers and the dependence on technicians. and reduce manpower. Therefore, it is important to develop an infrared image recognition method to help deal with large amounts of infrared image data. In the early stage, the detection based on infrared or infrared thermal was very limited by thermography and the level of computer calculations. Researchers focused on removing image noise with filter algorithms including Gaussian filter, adaptive median filter, median filter, etc. Also, the image segmentation technique, including object segmentation or fault region segmentation, is performed simultaneously, represented by threshold segmentation by the OTSU method[7-9]. Gray histogram methods [10], K-means [10,11], similarity of adjacent regions, growth algorithm based on area and morphological opening and closing operations [12], edge detection [13] and PCNN (pulse coupled neural network) pulse) [14] is presented. In [15], a new non-contact and non-intrusive method for monitoring electrical transformers and detecting their defects based on infrared thermography imaging techniques (IRT) imaging techniques is adopted. In [16], a method based on deep learning is proposed in transformer failure detection. In this paper, in addition to accuracy-based analysis, an in-depth evaluation is presented to show the most suitable architecture in thermal image classification. [17]proposes a fault detection model that includes adaptive synthetic oversampling (ADASYN), reconstructed data method, and an improved Mask Region convolutional neural network (CDCN). In [18], by building an object recognition system with a Convolutional Neural Network (CNN) frame, the Bush frame can be accurately extracted. To detect the fault area of bushings and the background, a pulse connected neural network based on simple linear iterative clustering is proposed to improve the fault area segmentation performance. In [19] a new interpretation of transformer failure analysis is proposed including new image features based on image processing technique. In [20], it has been proposed to detect faults in the power network using Gabor filters. In this method, the investigated image passes through a filter bank, then the output of the filter is thresholded. By combining the output of different filters, a suitable pattern of the defect can be obtained. In [21], in order to solve the key problem of automatic fault detection technology of power equipment, projected transformation has been introduced to extract the features of the thermal image in the system, using the characteristics of the equipment in this design. In [22]. they have proposed GLCM based equipment fault detection. In [23], a new transformer failure interpretation approach has been proposed to detect winding short circuit fault in transformer, radial deformation and bushing faults using polar diagram and have introduced digital image processing in which various unique image features of a polar plot are extracted using geometric dimensions, invariant moments. In [24], an online technique is introduced to detect internal faults in a power transformer by constructing a voltage-current (V-I) location diagram to provide the current status of the transformer's health status. However, the above methods are only applicable for images with a relatively simple background. For those infrared images with complex background, the main image segmentation methods have strong limitations. In order to overcome these complications, multi-resolution transformation methods are proposed[25]. Wavelet transform is one of the important mathematical multi-resolution transforms, which obtains a time-scale representation of the digital signal using digital filtering techniques [26]. In this transformation, it will be calculated by changing the scale of the analysis window, changing the location of the window in time, multiplying it by the signal and integrating it in all times. In the discrete case, filters of different cutoff frequencies are used to decompose the signal at different scales[27-29]. The signal is passed through a series of high-pass filters to analyze high frequencies and through a series of low-pass filters to analyze low frequencies. The degree of signal resolution, which is a measure of the amount of detail in the signal, is varied by filtering operations and scaled by upsampling and Subsampling (downsampling)[30].Bandelet are one of the multi-resolution transforms that have been developed to overcome the challenges and shortcomings of the wavelet transform. Bandelet calculates a geometric flow of an image to better extract smooth edge information. Bandelet decomposition is applied to the orthogonal wavelet coefficients or wavelet filter bank of an image and is calculated by geometric orthogonal transformation through orthogonal filters. As a result of this process, a different transformation is obtained from each geometric direction and they can be processed to find the optimal set of filters with the best basic algorithm. The characteristics of wavelets can be manipulated by selecting low-pass and high-pass square mirror filters [31].Morphological component analysis (MCA) can decompose an image into two components[32]. The advantages of MCA include completeness and effectiveness. (1) Completeness: In multiscale transformation, transformation bases are built to reveal salient features of an image. (2) Effectiveness: The sparse representation uses a complete dictionary with more columns than rows and models the image of the 1-D representation as a linear combination of

columns (atoms) [33]. The stability and reliability of a power system depends in many ways on the condition and health of power transformers [29]. As one of the most expensive and important elements, the power transformer is a very necessary element whose failure and damage may cause the power system to be interrupted. Therefore, it seems that transformers should be constantly inspected and reviewed in order to apply preventive repairs. In addition to the cost and the need for experienced manpower, visual inspection for power transformers is not very accurate and is always associated with errors. For this purpose, using methods based on image processing can be helpful [34]. Thermal or infrared IR cameras that work with infrared technology are an efficient way to provide preventive maintenance in power transformers and failure detection along with efficient image processing methods. It seems that it is necessary to propose a new and efficient method in diagnosing the failure of power transformers [35]. Therefore, in this research, a method based on the morphological component analysis will be presented in order to detect faults in power transformers. In the proposed method, taking advantage of the morphological diversity of images and the advantages of MCA, the component (carton, texture, and edge) is considered as a feature and will be entered into a classification for failure detection. The innovations of this research can be stated as follows:

- Improving the accuracy of transformer failure detection based on the analysis of morphological components

In the following, this article is divided as follows. In the second part, the principles of image processing in detecting defects in infrared images are presented. In the third part, the suggestion method will be presented. In the fourth part, the evaluation of the proposed method will be done. Finally, the conclusion of the article will be presented in the fifth section.

## 11. MATERIAL AND METHOD

The passage of electric current through various circuits and components in a device, devices and electrical equipment is always associated with heat production, it seems that by measuring the temperature and preparing thermal photographs of electrical components and equipment, it is a reliable guide in determining the weak points that may be in The future will lead to major connections. Hot spots in electrical equipment are often caused by not being strong, oxidation or corrosion of connections, as well as asymmetry of phases or insulation failure of coils. By using IR infrared image recording equipment in measuring the temperature of equipment such as power transmission lines, high voltage substations, transformers, switches, fuses, cables and all control equipment and electrical panels, it reveals them before they lead to destructive events. It was solved in the electrical system [36]. Unique information is extracted from infrared images related to electric energy distribution network equipment and used to show the performance status of this equipment in electric energy distribution smart networks. In addition to morphological features that can be observed, features such as wavelet features and statistical features are also effective in diagnosis. The method of analyzing infrared images in electrical energy distribution network equipment based on machine vision and image processing, data pre-processing, dimension extraction and reduction, classification and application are discussed [37].

### 11.1. Wavelet Transform

Wavelet transformation is one of the mathematical transformations in multi-precision domains. This conversion is a desirable strategy for establishing an optimal balance between time accuracy and frequency accuracy [38]. At higher frequencies, the wavelet transform gains time-domain information at the cost of losing frequency-related information. While at lower frequencies, it gains frequency information at the expense of temporal information loss. As the Fourier transform is defined based on an integral, the wavelet transform can also be defined based on an integral as follows:

$$W_{X(s,u)} = \int_{-\infty}^{+\infty} X(t) \Psi_{s,u}(t) dt \quad (1)$$

In the above integral, the input signal  $x(t)$  is related to the wavelet by means of the transfer parameter  $u$  and the coherence parameter  $s$ . This transform transforms a signal into coefficients that represent time-frequency information [39]. These coefficients have more time accuracy at high frequencies and more frequency accuracy at low frequencies. The homogeneity parameter enables the wavelet to exchange information in frequency events [40].

The DWT method overcomes the drawbacks and weaknesses of the fast Fourier transform (FFT) [41]. Signal analysis based on FFT works well when its frequency spectrum is not dependent on time, in other words, it is statistically stationary, but many signals are non-stationary in nature. Time-

frequency representation of time series signals is an attractive way to capture frequency information at low frequencies and time information at high frequencies. The DWT method is one of the techniques based on multi-resolution analysis [28]. For the signal X(t), the wavelet transform is defined as follows:

$$WT_X(j, k) = \frac{1}{\sqrt{a_0^j}} \int x(t) \Psi^* \left( \frac{t - ka_0^j b_0}{a_0^j} \right) dt \tag{2}$$

In equation (2), x(t) and Ψ are the initial signal and the wavelet transform function, respectively, and also the transmission parameter and a<sub>0</sub><sup>j</sup> is the scale parameter. where j represents the wavelength. Usually a<sub>0</sub> = 2, so scales are sampled during a binary sequence. The sample time domain is specified using Kb<sub>0</sub>. At each level of analysis, approximation coefficients CA and detail coefficients CD are created by passing the X(n) signal through the high-pass H and low-pass L filters [29]. Approximation coefficients and detail coefficients are obtained by the following equations [37]:

$$CA_j = \sum_{n=-\infty}^{\infty} X_{j-1}(n)l(n - 2k) \tag{3}$$

$$CD_j(k) = \sum_{n=-\infty}^{\infty} X(n)l(n - 2k) \tag{4}$$

Wavelet transform may not be the best choice for edge detection in natural images. This observation is based on the fact that wavelets are blind to the smoothness of edges commonly found in images. Therefore, there should be a new multi-resolution approach that is more flexible and efficient in capturing edge smoothness of images, should be used in edge detection and image segmentation applications.

The bandlet approach is a multi-resolution approach in the field of curve transformation. Their reasoning is as follows:

- This approach successfully removes noise, because the noise is not part of the structural information of the image, and the curve transform does not generate coefficients for the noise.
- But the transformation of the curve is defined in polar coordinates, which makes it difficult to translate it into Cartesian coordinates.

For segmentation in an image and especially for finding image edges for infrared images in transformers, they have limitations in determining wavelet base defects, because they are not well adapted to detect highly anisotropic elements such as alignments in an image. Bandlet transform performs better than wavelet transform in displaying salient image features such as edges, lines, curves and contours due to its anisotropy and directionality properties. Therefore, it is suitable for multi-scale edge-based color image enhancement.

Bandlet (Contourlet) conversion consists of two steps:

Subband decomposition and directional transformation .

A Laplacian pyramid (LP) is first used to capture point discontinuities in the subband decomposition step, then directional filter banks are used to link point discontinuities to linear structures. The overall result is an expansion of the image using basic elements such as contour segments and curves in the image and is named bandlet transform. Figure 1 shows the bandlet conversion flow diagram. The image is first decomposed into subbands by LP transformation, and then, each detail image is analyzed by directional filter banks (DFB).

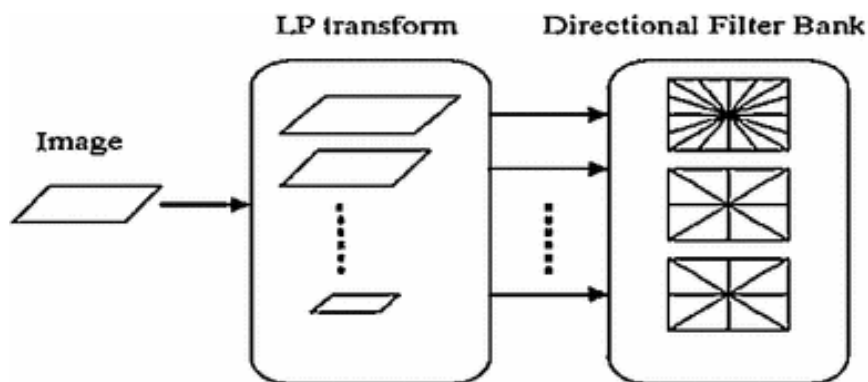




Fig. 1. Bandlet conversion performance.

In this section, the image separation method based on MCA Morphological component decomposition is described. The first related issue is the dimensions of the original data set, especially in far-infrared images with low and even appropriate spectral resolution such as hyperspectral data, and in order to reduce the dimensions of the data, this issue is very important. Dimensionality reduction of remote image classification can be done using all the original image information.

MCA-based image segmentation is a method that allows us to separate features in an image when these features present different morphological aspects. For an image  $y \in R^N$ , where N is the number of pixels in y, and a given texture feature, the task is to find the most sparse solution to the problem as follows:

$$\begin{aligned} x &= \operatorname{argmin} \|x\|_1 \\ \text{subject to: } y &= Ax \end{aligned} \tag{5}$$

where  $x \in R^K$  denotes the sparse MC coefficients,  $A \in R^{N \times K}$  denotes the associated dictionary, and K denotes the number of atoms in the dictionary (usually  $K > N$ ). which is a portable convex constrained optimization problem and can reduce the computational complexity by linear programming. In image texture separation, we often use MCA to decompose an image into texture and content components. In the work presented in [9], MCA-based decomposition is used to classify hyperspectral images, where the obtained results were very promising. For a given image y with N pixels, the goal of MCA is to split it into two components: a smoothness component  $y_s$  and a texture component  $y_t$ , respectively. These components represent the original image under a linear combination as follows. where n is the remainder in the image approximation.

$$y = y_s + y_t + n \tag{6}$$

## 12. PROPOSED METHOD

The main goal of this research is to detect transformer failures in infrared images using MCA principal components analysis. After improving the image quality with the help of pre-processing methods, MCA principal component analysis will be applied to the image. To classify and identify the failure, the relevant features of the cartoon texture and edge will be used for classification. The general block diagram of the proposed method is shown in Fig. 2.

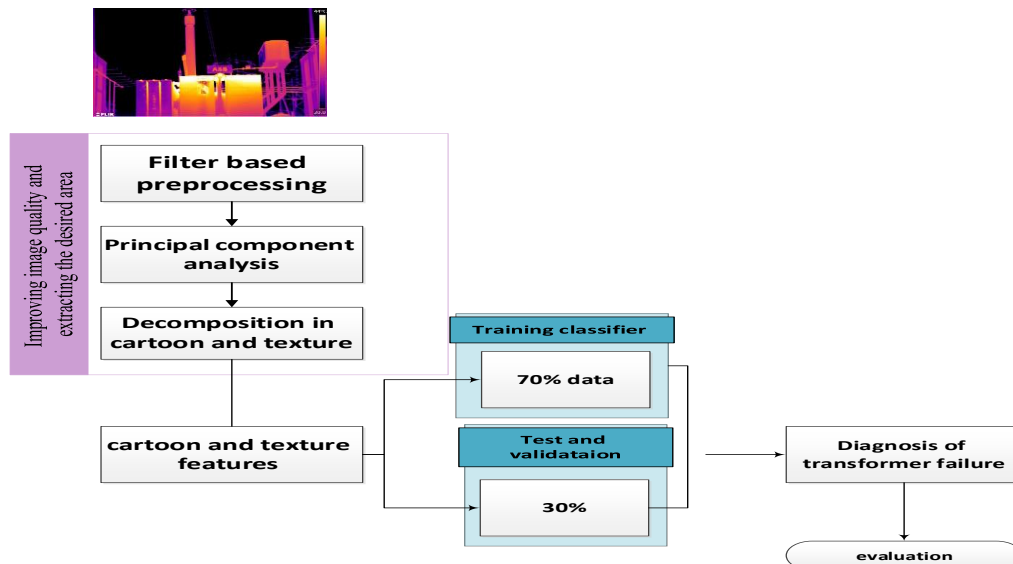


Fig. 2. Block diagram of the proposed method.

### The first step: improving the quality of the infrared image

Infrared images consist of a gray color spectrum that lies between white and black. Sometimes, due to the low contrast in the image, it is very difficult to distinguish the difference between adjacent pixels that have very close colors. Since the defect in an equipment discussed in this research may be similar to the rest of the parts, it is necessary to improve the quality of the image in such a way that it becomes easier to detect this defect in the transformer [17]. Bidirectional filtering is used as an edge preservation tool in image enhancement applications [4]. Along with a low-pass spatial kernel (which helps with smoothing), it uses a kernel to prevent smoothing near the edges. As a result, the filter is able to smooth homogeneous areas and preserve sharp edges at the same time. It was shown that spatial and domain kernels are typically Gaussian, which can be improved by adjusting the width and center of the Gaussian domain kernel at each pixel, and the enhancement capacity of the two-way filter.



Fig. 3. (a) The original image and (b) the improved rendering based on filtering in the proposed method.

### The second step: applying MCA principal components analysis

Morphological component analysis (MCA) is a new method that allows us to separate features in an image when these features present different morphological aspects. We show that MCA can be very useful for decomposing images into textures, edges and piecewise smooth (cartoon) or for feature extraction applications. To classify the defects in the infrared images using the neural network method or any other classifier in machine learning, first the features based on the morphological component analysis (MCA) are extracted from the infrared image and by them the neural network or another classifier. It is taught in machine learning.

In the framework of thin representation, a dictionary  $U = [\phi_1, \dots, \phi_T]$  is viewed as an  $N \times T$  matrix. When  $T > N$  or even  $T \gg N$ , the dictionary is too complete and is built by merging several dictionaries. An image  $x \in \mathbb{R}^N$  (an image with  $N$  pixels can be expressed as a lexically ordered 1-D vector) as a linear combination of elementary atoms  $M$  ( $M < T$ ) of the dictionary, according to the model equation has been made

$$\min_{\alpha \in \mathbb{R}^m} \|\alpha\|_0 \quad s.t. \quad x = \phi\alpha. \quad (7)$$

$$x = \phi\alpha = \sum_{i \in I_m} \alpha[i] \phi_i \quad (8)$$

where  $\alpha[i]$  is the representation coefficients of  $x$ .  $I_m$  is a subset of  $[1, T]$  and  $\text{Card}(I_m) = M$ .  $\phi_i$  represents  $U$  atoms. Obviously, from Eq. (7),  $x$  has a large number of candidate representations. It is the thinnest purpose of representation. Therefore, the sparse representation problem requires the following minimization solution.

$$\min_{\alpha \in \mathbb{R}^m} \|\alpha\|_1 \quad s.t. \quad x = \phi\alpha. \quad (9)$$

The complexity of the problem is formulated in equation (9). grows exponentially with the number of dictionary columns because the problem is non-convex. To reduce the complexity, the non-convex  $l_0$  scattering criterion is replaced by the  $l_1$ -norm [38]. Therefore, equation (9) becomes a portable convex optimization problem that can be solved by the base tracking (BP) basis pursuit [39].

In order to extract cartoon and texture features in MCA, the infrared images are systematically decomposed into texture and cartoon parts. The coefficients created in the texture and cartoon layers are tight enough to be used for the feature. If  $\alpha_c$  and  $\alpha_t$  represent the thin representation of texture and cartoon images respectively, the feature vector resulting from each visualization is  $[\alpha_c \quad \alpha_t]$ . BCR simplification based on block coordinates is used to calculate  $\alpha_c$  and  $\alpha_t$  pseudo-digram blocks.

### Bandlet features

The bandlet transform, which is a combination of the previous two transforms, allows us to analyze the image with different block sizes. The work process is that first, the image is decomposed into a set of wavelet bands and the analysis of each band will be by Reglet transformation. The size of the blocks can be changed in each level. In fact, it is a two-dimensional transformation that cannot be separated into one-dimensional transformations parallel to the coordinate axes. A bandlet transform is presented to optimally represent two-dimensional discontinuities. In this research, 2 levels of Corlott transformation are used. Figure (4) shows the filter banks resulting from bandlet transformation.

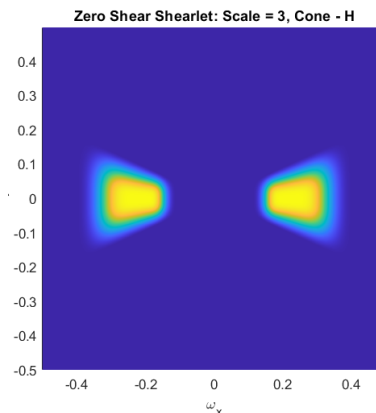


Fig. 4. Bandlet filterbanks

### Fourth step: error detection

In order to reduce the dimension and remove duplicate information, principal component analysis has been used in this research. The classification model is responsible for separating and categorizing based on a series of input data, and as a result, it finally achieves the goal of detecting failure in the desired equipment. Common machine learning algorithms including K nearest neighbor (KNN), decision tree and random forest, support vector machine (SVM), Artificial neural network (ANN) have been used in this research [44].

## 13. RESULTS

In this research, the proposed method will be evaluated with multiple evaluation criteria. The system for simulating the proposed method has hardware with 7 cores (Core™ (i7 CPU)) and a working frequency of 2.60GHz. The RAM available in this system is 16 GB. The Windows operating system installed on this device is version 11. The simulations were carried out in the used MATLAB 2020b software. The criteria of accuracy, sensitivity, positive predictive value and negative predictive value were used. Relationships are defined as follows. Necessary explanations regarding the values used in relations (10) to (12) have been stated[45].

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (10)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (11)$$

$$Specificity = \frac{TN}{TN+FP} \quad (12)$$

In above equation True Negative :TN, False Negative :FN, True Positive :TP, False Positive :FP. The desired criteria in the evaluation of the proposed method have been checked separately on the collected database. In order to determine the best parameters in the classification and faults and breakdowns in transformers, several tests were conducted on the SVM and K nearest neighbor KNN support vector machine classifications, as well as the random forest. Based on the tests, the best parameters have been announced as follows:

- In SVM window Radial based function RBF
- In the KNN classification, Euclidean metric criteria and neighborhood radius 3

- 1000 trees in the forest

In the whole classification process, 70% of the data is used for training and 30% for testing.

### 13.1. Research database

Saman Niro Sepahan Company is one of the active companies in the field of troubleshooting and fault detection in power networks, which uses Thermovision thermal cameras to identify faults. The image database used in this research is images collected by Saman Niro Sepahan Technical and Engineering Company. The used camera model is Trotec AC080V made in Germany. The number of images in this research is 300 images of different equipments including transformers, insulators, arresters and other equipments.

### 13.2. Quality evaluation

An example of qualitative fault diagnosis is shown in this article. Figure 5 shows the main image of the evaluated database. The results of Figure 6 show that the proposed method has achieved acceptable results. The quality of this method can be discussed from two perspectives. First, the algorithms based on morphological component analysis are among the most powerful zoning methods and have many defects in improving zoning. Second, the use of pre-processing methods has been able to reduce the system error to a great extent.

### 13.3. Quantitative assessment

Table 2 shows the fault results in transformer equipment without dimension reduction and with features based on wavelet transformation. As can be seen from Table 2, the detection results are not very promising and it seems that to improve the results, a suitable dimension reduction method and a suitable feature descriptor are needed in order to avoid overfitting and improve the results. The results of Table 1 show the superiority of the KNN classification method in the defect in the transformer equipment in the state without dimension reduction.



Fig. 5. The main image of existence in the database.

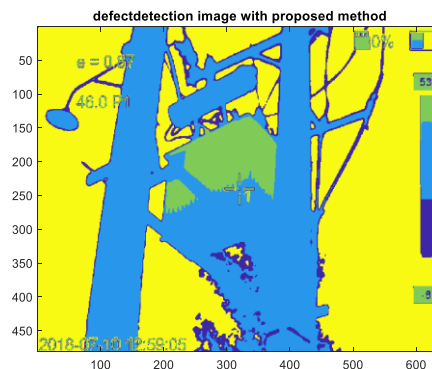


Fig. 6. Zoning and error detection in the proposed method.

### The results of PCA classification

Support vector machine, random forest and K-nearest neighbor classifiers were used to classify the infrared image data in the database. For classification, first pre-processing was done on the desired

database images. After feature extraction with the help of the proposed method, the next step of reduction was done with the help of PCA, then the classification was done with the help of the introduced classes. Table 2 shows the results of this classification. As can be seen from the results of Table 3, the results obtained are in a similar range. However, the KNN classifier has been able to perform as the best classifier with the best calculated accuracy value as well as lower sensitivity, higher accuracy, and although lower recall rate. The random forest classifier as well as the SVM classifier have similar results, and in any case, the accuracy of the results has not reached 90%.

#### The results of the classification of the proposed method

The given dimensionality reduction features are classified using the proposed method based on main morphological components as well as multi-resolution analysis with KNN classifier and SVM support vector machine as well as random forest classifier. The results of the proposed method in Table 4 show the results of this classification.

**Table 1.** Simulation results without dimension reduction

Evaluation criteria	K nearest neighbor(%)	Support vector machine (%)	Random forest(%)
Precision	75.26	68.13	71.76
Sensitivity	71.38	66.75	72.72
Accuracy	72.33	63.75	70.27
Recall	67.21	71.89	69.43
Criterion F	87.01	80.01	74.00

**Table 2.** The results of the simulation with dimension reduction with the help of PCA.

Evaluation criteria	K nearest neighbor(%)	Support vector machine(%)	Random forest (%)
Precision	86.26	79.13	75.64
Sensitivity	81.38	80.57	66.72
Accuracy	89.33	83.75	76.26
Recall	77.21	85.89	75.43
Criterion F	87.25	80.80	74.02

**Table 3.** Simulation results along with dimension reduction in the proposed method.

Evaluation criteria	K nearest neighbor(%)	Support vector machine(%)	Random forest (%)
Precision	95.81	94.10	98.18
Sensitivity	91.51	85.37	86.25
Accuracy	84.54	81.14	63.38
Recall	91.25	90.87	97.87
Criterion F	94.28	91.23	99.25

As it is clear from the results of Table 4, the RF classifier with K nearest neighbor has been able to get better results. It should be noted that the accuracy and sensitivity values obtained and the recall rate are also acceptable results. As it can be seen from the graph, the combination of the extracted features has led to better results, and this proves the important sub-hypothesis of this research, including the improvement of results in classifications by reducing the dimension with the help of the proposed method. This superiority is still perceptible in all the criteria used. The F criterion is also included in these evaluations. In this case, the random forest classifier has been able to obtain the best result with 0.99 in the accuracy criterion as well as the F criterion with one hundred features.



### 13.4. Comparison with other studies

The proposed method in this article is compared with other related researches in fault diagnosis in transformer equipment. These methods include RNN recurrent neural networks [32] and deep learning [33]. In recent years, methods based on deep learning are among the most common methods and of course with good efficiency. Various researches have used methods based on deep learning. Although these methods are acceptable and highly accurate, they require a large database for simulation. The desired methods for comparison have been simulated on the database of this research. Therefore, a correct comparison and correct evaluation has been made. The results of the comparison of the proposed method with the desired researches are shown in table (5). As can be seen from the results of this table, the proposed method has better accuracy than other methods based on deep learning. The database used in all these researches is the database collected in this research by Saman Niro Sepahan Company.

**Table 4.** Comparison of the proposed method with other articles.

The presented model	Research	Average accuracy
2 D-CNN	[32]	94.10
RNN	[33]	95.91
Proposed	-	99.00

## 14. CONCLUSION

In this research, the relationship between In this article, the proposed method for detecting defects in transformer equipment was evaluated in infrared images. In the proposed method, several features in the time domain including morphological features (edge and cartoon), multi-resolution features based on wavelet transform with Dabich's filter bank were also extracted. A method based on PCA principal component analysis was introduced. The feature vector matrix was classified by support vector machine, random forest and k nearest neighbor classifiers. The parameters of recall rate, precision, accuracy, specificity and f-criterion are evaluated in three categories: SVM, KNN and RF. The RF classification has the best result with numerical values higher than 99% in the proposed method in feature selection based on the PCA algorithm.

## REFERENCES

- [1] X. Ouyang, Q. Zhou, H. Shang, Y. Zheng, S. Pan, and J. Luo, "Towards a Comprehensive Evaluation on the Online Methods for Monitoring Transformer Turn-to-Turn Faults," *IEEE Transactions on Industrial Electronics*, 2022.
- [2] L. Raeisian, H. Niazmand, E. Ebrahimnia-Bajestan, and P. Werle, "Thermal management of a distribution transformer: An optimization study of the cooling system using CFD and response surface methodology," *International Journal of Electrical Power & Energy Systems*, vol. 104, pp. 443-455, 2019.
- [3] E. Karimi, and A. Ebrahimi, "Considering risk of cascading line outages in transmission expansion planning by benefit/cost analysis," *International Journal of Electrical Power & Energy Systems*, vol. 78, pp. 480-488, 2016.
- [4] E. Karimi, A. Ebrahimi, and M. R. Tavakoli, "How optimal PMU placement can mitigate cascading outages blackouts?," *International Transactions on Electrical Energy Systems*, vol. 29, no. 6, pp. e12015, 2019.
- [5] E. Karimi, and A. Ebrahimi, "Probabilistic transmission expansion planning considering risk of cascading transmission line failures," *International Transactions on Electrical Energy Systems*, vol. 25, no. 10, pp. 2547-2561, 2015.
- [6] M. Emadi, Z. Jafarian Dehkordi, and M. Iranpour Mobarakeh, "Improving the Accuracy of Brain Tumor Identification in Magnetic Resonance using Super-pixel and Fast Primal Dual Algorithm," *International Journal of Engineering*, vol. 36, no. 3, pp. 505-512, 2023.
- [7] F. B. Zade, and M. Emadi, "The Improvement of Breast Cancer Diagnosis Rate in Magnetic Resonance Imaging (MRI) using Fusion of Super Pixels and Fuzzy Connectedness," *Majlesi Journal of Telecommunication Devices*, vol. 10, no. 4, pp. 137-145, 2021.
- [8] S. A. M. Mobarakeh, and M. Emadi, "Improvement of the Identification Rate using Finger Veins based on the Enhanced Maximum Curvature Method using Morphological Operators," *Majlesi Journal of Telecommunication Devices*, vol. 11, no. 1, pp. 1-8, 2022.
- [9] M. Soleimani, and J. Coykendall, "Divisors and Factorization Type."
- [10] R. S. Marques, A. Conci, M. G. Perez, V. H. Andaluz, and T. M. Mejia, "An approach for automatic segmentation of thermal imaging in Computer Aided Diagnosis," *IEEE Latin America Transactions*, vol. 14, no. 4, pp. 1856-1865, 2016.
- [11] N. Liu, J. Zhang, and X. Wu, "Asset analysis of risk assessment for iec 61850-based power control systems—part i: methodology," *IEEE Transactions on Power Delivery*, vol. 26, no. 2, pp. 869-875, 2010.

- [12] A.-V. Vo, L. Truong-Hong, D. F. Laefer, and M. Bertolotto, "Octree-based region growing for point cloud segmentation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 104, pp. 88-100, 2015.
- [13] A. Aslam, E. Khan, and M. S. Beg, "Improved edge detection algorithm for brain tumor segmentation," *Procedia Computer Science*, vol. 58, pp. 430-437, 2015.
- [14] D. Zhou, and Y. Shao, "Region growing for image segmentation using an extended PCNN model," *IET Image Processing*, vol. 12, no. 5, pp. 729-737, 2018.
- [15] A. Mahami, T. Bettahar, C. Rahmoune, F. Amrane, M. Touati, and D. Benazzouz, "A New Transformer Condition Monitoring Based on Infrared Thermography Imaging and Machine Learning," *Advanced Computational Techniques for Renewable Energy Systems*, pp. 408-418: Springer, 2023.
- [16] G. Sakalli, and H. Koyuncu, "Identification of asynchronous motor and transformer situations in thermal images by utilizing transfer learning-based deep learning architectures," *Measurement*, vol. 207, pp. 112380, 2023.
- [17] Z. Li, Y. He, Z. Xing, and J. Duan, "Transformer fault diagnosis based on improved deep coupled dense convolutional neural network," *Electric Power Systems Research*, vol. 209, pp. 107969, 2022.
- [18] J. Jiang, Y. Bie, J. Li, X. Yang, G. Ma, Y. Lu, and C. Zhang, "Fault diagnosis of the bushing infrared images based on mask R-CNN and improved PCNN joint algorithm," *High voltage*, vol. 6, no. 1, pp. 116-124, 2021.
- [19] L. Zhou, J. Jiang, X. Zhou, Z. Wu, T. Lin, and D. Wang, "Detection of transformer winding faults using FRA and image features," *IET Electric Power Applications*, vol. 14, no. 6, pp. 972-980, 2020.
- [20] L. Zhu-Mao, L. Qing, J. Tao, L. Yong-Xin, H. Yu, and B. Yang, "Research on thermal fault detection technology of power equipment based on infrared image analysis." pp. 2567-2571.
- [21] S. Anoop, K. Ilango, A. Dhieep, C. Thomas, J. Jose, A. Kumar, and P. Ajith, "Thermal stress monitoring and pre-fault detection system in power transformers using fibre optic technology." pp. 886-891.
- [22] H. Shen, L. Zhu, X. Hong, and W. Chang, "ROI extraction method of infrared thermal image based on GLCM characteristic imitate gradient." pp. 192-205.
- [23] O. Aljohani, and A. Abu-Siada, "Application of digital image processing to detect transformer bushing faults and oil degradation using FRA polar plot signature," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 24, no. 1, pp. 428-436, 2017.
- [24] A. Abu-Siada, and S. Islam, "Image processing-based on-line technique to detect power transformer winding faults." pp. 5549-5554.
- [25] M. Soltani, M. Bahadori, and M. Soleimani, "Optimal Predictive Maintenance and Spare Part Inventory Policies for a Degrading System Subjected to Imperfect Actions," *Available at SSRN 4558570*, 2023.
- [26] A. Mahami, C. Rahmoune, M. Zair, T. Bettahar, and D. Benazzouz, "Automated Transformer fault diagnosis using infrared thermography imaging, GIST and machine learning technique," *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, vol. 236, no. 4, pp. 1747-1757, 2022.
- [27] S. Aryanmehr, M. Karimi, and F. Z. Boroujeni, "CVBL IRIS Gender Classification Database Image Processing and Biometric Research, Computer Vision and Biometric Laboratory (CVBL)." pp. 433-438.
- [28] M. Emadi, M. Karimi, and F. Davoudi, "A Review on Examination Methods of Types of Working Memory and Cerebral Cortex in EEG Signals," *Majlesi Journal of Telecommunication Devices*, vol. 12, no. 3, 2023.
- [29] M. Harouni, M. Karimi, A. Nasr, H. Mahmoudi, and Z. Arab Najafabadi, "Health Monitoring Methods in Heart Diseases Based on Data Mining Approach: A Directional Review," *Prognostic Models in Healthcare: AI and Statistical Approaches*, pp. 115-159: Springer, 2022.
- [30] M. Aqil, A. Jbari, and A. Bourouhou, "ECG signal denoising by discrete wavelet transform," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 13, no. 09, pp. 51-68, 2017.
- [31] A. R. Khan, F. Doosti, M. Karimi, M. Harouni, U. Tariq, S. M. Fati, and S. Ali Bahaj, "Authentication through gender classification from iris images using support vector machine," *Microscopy research and technique*, vol. 84, no. 11, pp. 2666-2676, 2021.
- [32] M. Karimi, M. Harouni, and S. Rafieipour, "Automated medical image analysis in digital mammography," *Artificial intelligence and internet of things*, pp. 85-116: CRC Press, 2021.
- [33] M. Harouni, M. Karimi, and S. Rafieipour, "Precise segmentation techniques in various medical images," *Artificial Intelligence and Internet of Things*, pp. 117-166, 2021.
- [34] M. Soleimani, F. Mahmudi, and M. H. Naderi, "On the Maximal Graph of a Commutative Ring," *Mathematics Interdisciplinary Research*, 2021.
- [35] M. Soleimani, and A. S. Mirshahzadeh, "Multi-class Classification of Imbalanced Intelligent Data using Deep Neural Network," *EAI Endorsed Transactions on AI and Robotics*, vol. 2, 2023.
- [36] M. Soleimani, F. Mahmudi, and M. Naderi, "Some results on the maximal graph of commutative rings," *Advanced Studies: Euro-Tbilisi Mathematical Journal*, vol. 16, no. suppl, pp. 21-26, 2023.
- [37] A. Rehman, M. Harouni, F. Zogh, T. Saba, M. Karimi, and G. Jeon, "Detection of Lung Tumors in CT Scan Images using Convolutional Neural Networks," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2023.
- [38] M. Karimi, M. Harouni, A. Nasr, and N. Tavakoli, "Automatic lung infection segmentation of covid-19 in CT scan images," *Intelligent Computing Applications for COVID-19*, pp. 235-253: CRC Press, 2021.
- [39] A. Raftarai, R. R. Mahounaki, M. Harouni, M. Karimi, and S. K. Olghoran, "Predictive models of hospital readmission rate using the improved AdaBoost in COVID-19," *Intelligent Computing Applications for COVID-19*, pp. 67-86: CRC Press, 2021.

- [40] V. Seena, and J. Yomas, "A review on feature extraction and denoising of ECG signal using wavelet transform." pp. 1-6.
- [41] A. Rehman, M. Harouni, M. Karimi, T. Saba, S. A. Bahaj, and M. J. Awan, "Microscopic retinal blood vessels detection and segmentation using support vector machine and K-nearest neighbors," *Microscopy research and technique*, vol. 85, no. 5, pp. 1899-1914, 2022.
- [42] L. Xie, Z. Li, Y. Zhou, Y. He, and J. Zhu, "Computational diagnostic techniques for electrocardiogram signal analysis," *Sensors*, vol. 20, no. 21, pp. 6318, 2020.
- [43] K. Zhang<sup>12</sup>, Y. Huang, C. Song, H. Wu, L. Wang, and S. M. Intelligence, "Kinship verification with deep convolutional neural networks," 2015.
- [44] T. G. Jan, and S. M. Khan, "A Systematic Review of Research Dimensions Towards Dyslexia Screening Using Machine Learning," *Journal of The Institution of Engineers (India): Series B*, pp. 1-12, 2023.
- [45] M. Soleimani, A. Harooni, N. Erfani, A. R. Khan, T. Saba, and S. A. Bahaj, "Classification of cancer types based on microRNA expression using a hybrid radial basis function and particle swarm optimization algorithm," *Microscopy Research and Technique*, 2024.

# Improving the Accuracy of Segmentation of Remote Sensing Images using Deep Learning

Adel Hamdy Dhayef<sup>1</sup>, Mehran Emadi<sup>2</sup> 

1- Department of Computer Engineering, Isfahan (Khorasgan) branch, Islamic Azad University Isfahan, Iran.  
Email: adeleng975@gmail.com

2- Department of Electrical Engineering, Mobarakeh Branch, Mobarakeh, Isfahan, Iran.  
Email: emadi.mehran49@gmail.com (Corresponding author)

## ABSTRACT:

Image segmentation is used to exploit remote sensing images with high resolution. The purpose of segmentation is to create different segments with common features. For example, residential areas, forests, rivers and other areas are obtained in subdivision. But the random position of different areas on the ground has caused the accuracy of segmentation methods to be low. Using deep learning methods can improve segmentation accuracy. In this paper, a SegNet convolution deep neural network is proposed for segmentation of high resolution (HR) remote sensing images. The proposed strategy is to improve the semantic segmentation performance of images. The proposed SegNets strategy is carried out in two steps. The proposed method has been evaluated with accuracy criteria and F1 score. The results show that the accuracy is improved by more than 4% compared to other methods based on deep learning. Also, other evaluation criteria such as ROC have been used. The results of this criterion also show the superiority of this proposed method.

**KEYWORDS:** Segmentation, Hyper Spectral Images, Remote Sensing, Morphological Operators, Convolution Neural Networks.

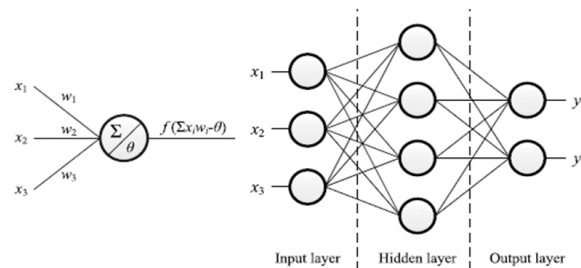
## 15. INTRODUCTION

In general, remote sensing (RS) is the technology of collecting information and taking pictures from the surface of the earth using aviation equipment such as airplanes, balloons or space equipment such as satellites [1, 2]. In other words, remote sensing is the science and art of obtaining information about any subject under investigation by a tool that is not in physical contact with it. The superior advantage of satellite information over other information sources is their repeated coverage of certain areas with a certain time interval. In remote sensing, information transmission is done using electromagnetic radiation (EMR) [3]. The data obtained from hyper spectral images in RS systems have been available to researchers since the early 1980s, and their use indicates the maturity of the technology. For the first time, the Hyperion probe was used by NASA in November 2000 to test the capability of an airborne vehicle. Airborne and space borne hyper spectral sensors (HS) sensors have many applications as one of the powerful and advanced tools in geology, agriculture and geography studies [4-7]. The use of this technology also began in the mid-80s, and the current advantages of remote sensing data and geographic information led to the development of this technology [8]. The preparation of hyper spectral imaging (HSI) is naturally more difficult and expensive than multispectral imaging (MSI) and this is due to the current advantages of these data (its high signal-to-noise ratio indicates the spectrum High quality data, as well as its spectral coverage and large number of channels, cause a very high spectral separation power, these data are generally a combination of 100 to 200 spectral bands (channels) with a thin band width between 5-10 nm, in While the data obtained from MS sensors are in the receiver of 5 to 10 channels with a relatively wider bandwidth (between 70-400 nm)[9]. (This technology is almost new and is used by researchers and scientists for mineral identification. plants, vegetation and artificial materials are used[10]). Identifying and distinguishing the existing components and categorizing hyper spectral images in order to use these images optimally and with higher efficiency is one of the important issues for the analysis and investigation of hyper spectral images[11]. Content analysis of satellite images is important in order to use these images in urban planning, agriculture, forestry, as well as management of underground resources and other applications[12]. Using satellite images, there is a greater need for automatic identification of land cover types, including roads, rivers, forests, and other natural and non-natural effects. which is one of the common methods of using

segmentation methods of these images [13]. Segmentation of satellite images is a hot and active topic of remote sensing due to the complexity of complications and the spatial and spectral heterogeneity of the urban and non-urban environment (man-made and natural), which has a large variety of artificial and natural surfaces [14-16]. Segmentation is widely used for mapping purposes. Accurate mapping of urban land use is used for various applications, including information on land use, urban management, change detection, urban planning and design, and environmental monitoring. For this reason, it is very important to have accurate and timely information about the status and changes in urban areas [17]. Various segmentation techniques have been presented for satellite images [18]. In the method presented in [19], the classification of hyper spectral images was done using the reduction of spatial spectral dimensions. This method is presented to overcome the challenges of a large number of spectral bands and data heterogeneity to improve classification. Classification of hyper spectral images was done with SVM and guided filter. In the method [20], considering that the combination of spatial features has been widely used, this method was used to extract spatial features in spectral-spatial using directed filter to classify hyper spectral images. A method for classifying hyper spectral images is presented in [21]. Researchers, in this research, presented a new spectral, spatial classification approach based on texture pattern separation for hyper spectral image classification. In [22], a method based on super pixels based on KNN nearest neighbor classifier is presented. In [23], a new method of local binary pattern (LBP) based on super pixel decision-making is proposed for the classification of hyper spectral images. In this method [24], open morphological transformations are used to separate bright (open) structures in the images, where bright means brighter than the surrounding features in the images. In [25], morphological summaries with subjective retrieval and guided MPs are first investigated for over-decision hyper spectral snapshot classification of city areas. Second, a supervised face extraction is developed to reduce the dimensionality of the morphological profiles created for prediction. In [26] proposed a joint hyper spectral and infrared image classification framework based on threshold-based local contain profile, where TLCP is a new design to suppress interference in spatial extractions. In [27], he used deep learning to quickly identify fires and predict possible spread for effective response in suppression. In [28], the most challenging problem of land use extraction in images with medium spatial resolution has been used using deep learning semantic segmentation. The random position, the completely random form of natural and human-made effects in satellite images has caused that the presented solutions cannot be effective in all applications. On the other hand, the existence of noise in these images is inevitable. In the existing methods for the analysis of hyper spectral images, uncertainty management has not been done properly. Uncertainty caused by noise, measurement error, and inaccurate content of images has a great impact on reducing the accuracy of identification and classification results of hyper spectral images [29]. It seems that it is necessary to provide useful, efficient and new methods in this field. Various methods have been proposed for image segmentation, which are mainly divided into five categories: thresholding, clustering, edge detection, region extraction, and feature extraction [30]. Since the existence of imprecise content in the images collected by remote sensing is unavoidable, it seems that the use of deep learning will be very efficient to remove the uncertainty and analyze the images [31]. Through the studies, it was found that although many studies and researches have been presented for the segmentation of satellite images, there is still a long way to go before reaching a favorable answer in segmentation with high accuracy. According to the studies conducted in the review of the research literature and the background of the research, it is clear that one of the most important challenges of this research is the similarity of the textures on the ground, the randomness of the position and shape of the textures and complications on the ground. The capabilities of fuzzy methods can be a way to overcome these challenges. Therefore, in this research, we have used a deep learning based method along with morphological operators to segment hyper spectral images. Deep learning can segment images with an acceptable accuracy by overcoming the limitations of uncertainty. In the following, this article is divided as follows. In the second part, deep learning is introduced. In the third part, the proposed method is presented with a block diagram. In the fourth part, the proposed method is evaluated. Finally, the conclusion of the article is presented in the fifth section.

## 16. DEEP NEURAL NETWORKS

Deep neural networks are multi-layered structures. These networks are designed to strengthen and remove the limitations of the two-layer neural network. They have two or more layered structures, which are called intermediate layers of the hidden layer. The change in the number of neurons in each layer and the number of hidden layers causes a change in the performance of the neural network. Next, a structure of a neural network is shown in figure (1).



**Fig. 1.** Structure of a perceptron multi-layer neural network.

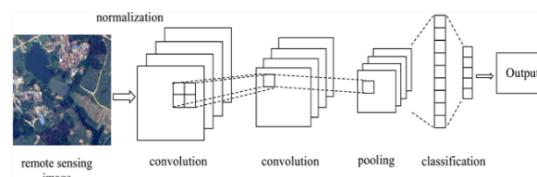
Since the output of each neuron is used as the input of the next neuron, the components of the neuron labeled  $j$  with the input of  $p_j(t)$  from the previous neurons include the following:

- The activation of the neuron state ( $a_j(t)$ ) depends on a selected location.
- If a function is constant, then a threshold is constant.
- The activation function whose goal is to activate a new one at a certain time ( $(t+1)$  of  $\theta_j$ ) with a net input  $p_j(t)$ .
- The activation function whose goal is to activate a new one at a certain time ( $(t+1)$  of  $\theta_j$ ) with a net input  $p_j(t) \circ_j(t+1) = f_{out}(a_j(t))$ .

It uses local for effective extraction of spatial information and joint weights to significantly reduce the number of parameters. Nowadays, in order to remote sensing image analysis, an unsupervised motion network has been suggested. To formulate a deep CNN model, unsupervised layer pre-training is used in this method. In compared to the unsupervised method, the supervised CNN may extract more effective features by using of class-specific information, which can be provided by the training examples [32].

### 16.1. Deep convolution neural network

One of the most popular deep learning models that specialize in spatial information discovery are Convolution neural networks. In this section, the working mechanism of CNN is briefly introduced. CNNs are widely used to discover latent spatial information in applications such as image recognition, ubiquity, and object search due to their salient features such as regular structure, good localization, and translation invariance. In BCI, in particular, CNN is supposed to capture distinct dependencies among patterns associated with different brain signals. A standard CNN architecture is shown in Figure 2. CNN consists of an input layer, two convolution layers with each integration layer, a fully connected layer and an output layer. A square patch in each layer shows the processing progress of a particular set of input values. The key to CNN is to reduce the input data in a way that is easier to recognize and with as little loss of information as possible.



**Fig. 2.** convolution neural network.

CNN has three dense layers: convolution layer, pooled layer and fully connected layer. The convolution layer is the main block of CNN, which consists of a series of filters for de convolution of the input data, followed by a non-linear transformation to extract geographic features. In implementing deep learning, there are several key parameters that need to be set in the convolution layer, such as the number of  $l$ , the size of each  $l$ , etc. The pooling layer generally follows the convolution layer. The purpose of the fusion layer is to gradually reduce the spatial size of the features. In this way, it can help reduce the number of parameters (eg, weight and basis) and computational burden. There are three types of integration operations: maximum, minimum, average. Take max integration as an example. The merge operation outputs the maximum value of the merge region in the result. Hyper parameters in the integration layer include integration operations, integration region size, steps, etc. In a fully connected layer, such as a basic neural network, nodes are fully connected to all previous activations. To extract spectral and spatial information of hyper spectral data simultaneously, it is reasonable to formulate 3D CNN. Furthermore, to solve the problem of over fitting caused by limited training samples of hyper spectral data, we design a regularization strategy, including modified linear unit (ReLU) and dropout



to achieve better generalization of the model. A complete CNN stage consists of a convolution layer and an aggregation layer. A deep CNN is built by stacking multiple convolution layers and combining the layers to create a deep architecture. First, the convolution layer is introduced. The value of neuron  $v_{ij}^x$  at position  $x$  of the  $j$ th feature map in the  $i$ th layer is shown below.

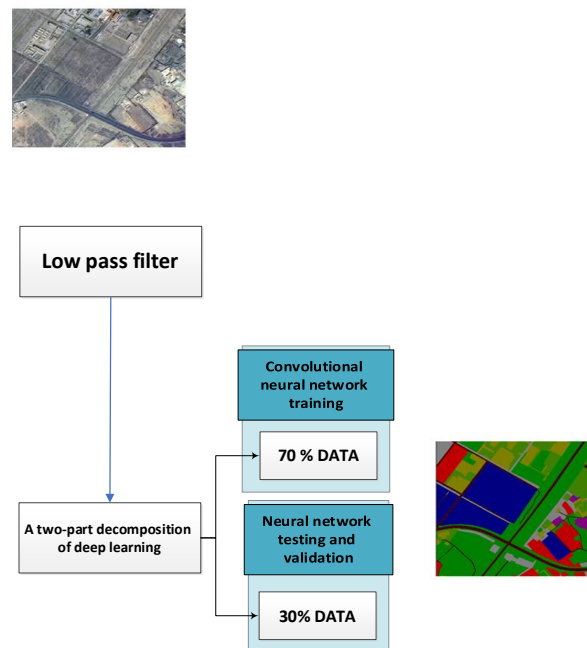
$$v_{ij}^x = g(b_{ig}) + \sum_m \sum_{p=0}^{p_i-1} w_{ijm}^p v_{(i-1)m}^{x+p} \quad (1)$$

First, the convolution layer is introduced. The value of neuron  $v_{ij}^x$  at position  $x$  of the  $j$ th feature map in the  $i$ th layer is shown below.

- $m$  lists the feature map in the previous layer (the  $(i - 1)$ th layer) connected to the current feature map.
- $w_{ijm}^p$  weight of position  $p$  connected to  $m$  feature map
- $P_i$  is the width of the kernel towards the spectral dimensions and the bias selection of the  $j$ th feature map in the  $i$ th layer.

## 17. PROPOSED METHOD

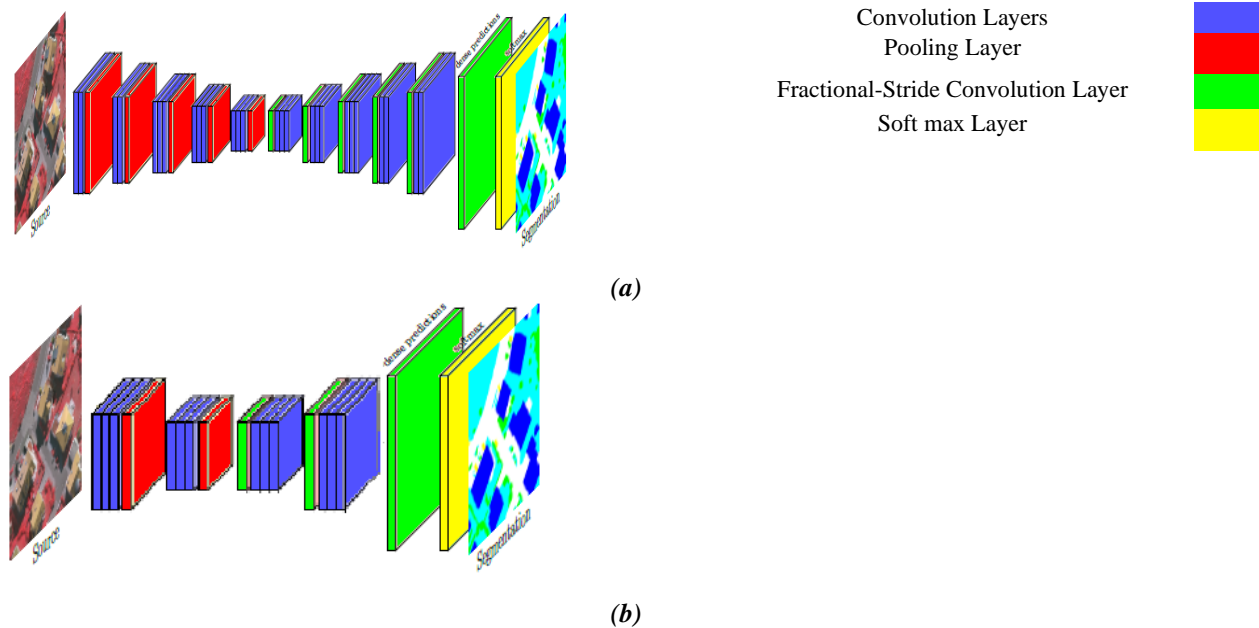
The main goal of this research is segmentation of hyper spectral images using deep learning method. In this regard, the quality of the image should be improved with the help of pre-processing methods. Then, for classification, segmentation should be done with the help of the proposed neural network based on convolution neural network. Figure 3 shows the block diagram of the suggested method. In the following, the suggested approach will be presented.



**Fig. 3.** Block diagram of the proposed method.

Deep neural networks have been widely used in the analysis of satellite images. Convolution neural networks are a type of deep neural networks that have shown their superiority in the segmentation of satellite images. CNN has been successfully used to segment these images. Although CNNs have shown their superiority, one of the major weaknesses of these methods is the need for a large historical database for training. Another is a weak generalization. SegNet is a type of CNN specially structured for semantic segmentation of images. But the commonly designed SegNet has only one mode for image segmentation, it can reduce the segmentation accuracy especially in small areas. In this paper, two-stage SegNet is proposed for HR remote sensing image segmentation. SegNet and SegNet-Basic as a type of CNN are shown schematically in Figure (4). SegNet is proposed to further address the issue of incomplete boundary delineation observed in other FCNNs. Perhaps the most important highlight of SegNet is that it requires significantly less memory than others. So when large-scale processing is needed, a larger area can be processed faster. In addition, the encryption part of SegNet is identical to that of popular classification networks such as VGG Net. Therefore, the benefits of pre-training can be achieved by initialization. SegNet has an encoder-decoder convolution architecture. Each encoder

consists of one or more convolution layers with batch normalization and nonlinear ReLU. Sampling in the decoder is done using max-pooling indices in the encoding sequence. The encoder is based on 13 convolution layers of the VGG-16 network followed by 13 corresponding decoders. SegNet-Basic is a smaller counterpart of SegNet, with only four layers each for encoder and decoder with a fixed feature size of 64. This model is trained using stochastic descent. The SegNet-Basic architecture is shown schematically in Figure 4, where the blue layers represent the convolution layers, including the ReLU and batch-normalization layers. Red layers indicate maximum accumulation. The green layer represents the pitch fractional tensional layer and the yellow layer is the Soft Max layer.



**Fig. 4.** Schematic of (a) SegNet, (b) SegNet-Basic architectures.

The steps of the proposed method are explained in the following steps:

- i. In the first step, the original image is applied to the grid to divide it into two classes. In the convolution neural network section, six convolution stages and ReLU activation function, six integration stages and three sampling stages are used. In this part, the image enters the convolution neural network and the output of the specified area is black and white binary images, in other words, the output of this network is shown in number 6. Black pixels in number 7 correspond to the first class, white pixels correspond to the second class. Dividing into two classes produces sharper edges compared to dividing into six classes at once. As explained, the selection of three subclasses of the first and second class is practically investigated. Based on extensive practical testing, building, tree and vegetation pixels are first class. And so, the road, the car, and the mess are left for second class. SegNet-Basic weights were used for more accurate segmentation in the next mode.
- ii. In the second step, the first class is divided into three subclasses by another network. In this step,  $32 \times 32$  packets are sampled from the local neighborhood in the output database of satellite images in the previous step. The sample images are fed to SegNet-Basic, which has  $3 \times 3$  filters and three dense layers with weights of 256, 128, and 2. There are also two Soft max because there is a fixed feature in the transfer in pool layers, which are also pool layers. used in this research. which has a batch size of 128 and cross entropy function. 1- The ReLU activity function is used to initialize the weights, which have a Gaussian distribution with zero mean. The training rate is also selected with an initial value of 0.0001. In order to avoid overtraining, this description has been increased to 0.3. The maximum number of courses is 1000. The rule with the least error is considered as the source model. This network will be trained with patch model. Also, for the final segmentation, the density layers are changed to their convolution equivalent and the loss function and regularization method are used to update the rules and weights. The output of this step, as shown in Figure 8, contains three tags: building, tree, low vegetation and some black pixels. Here, the black pixels belong to the second class, and the network does not decide which subclasses they belong to, i.e., the unknown pixels are black.
- iii. The third step divides the pixels of the second class that are labeled into three subclasses.

## 18. EVALUATION

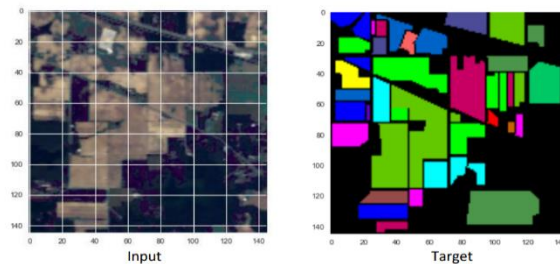
Segmentation and zoning of hyper spectral images in remote and satellite sensing has always faced many challenges. The random shape and position of land features, including man-made or natural features, the low contrast of these images has caused several methods to be presented to identify this important issue. In this research, an efficient and at the same time simple method based on two-stage SegNet deep learning is presented. In this article, the accuracy criterion is used to evaluate the proposed segmentation. Relationship 5-4 shows this important.

$$\text{Accuracy} = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{FP} + N_{TN} + N_{FN}} \quad (2)$$

In these equations, NTP is the number of positive correct nesses, NTN is the number of negative correct nesses, NFP is the number of positive errors and NFN is the number of negative errors in detecting the pixel type in the input samples.

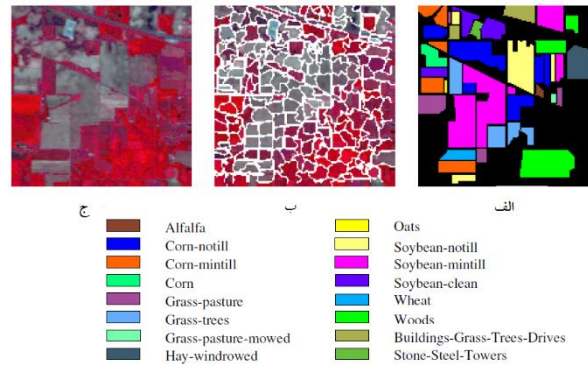
### 4.1. Database Used

To analyze and check the designed algorithm and compare the results of the proposed method, a set of other standard images have been used for evaluation. The images used in this research belong to the Indian Pines Satellite database, which contains a large collection of cloud spectrum images from different locations, which are prepared with different sizes and qualities. Different phenomena and covers can be seen in the prepared images. Urban texture, green space, river, sea, forest, mountain, etc. are among the textures that can be recognized by the designed algorithm in these images. The Indian Pines database is one of the well-known standard image sets in the field of hyper spectral image processing. An example of the segmentation results obtained from the implementation of the proposed algorithm in this research on the Indian Pines database is shown in Figure 5. The images of this standard database are used for better recognition of extractable areas of different colors in this image.



**Fig. 5.** Segmentation of a sample image from the Indian Pines database.

As mentioned earlier, the region extraction algorithm does not work properly in some parts of the image and the detected regions for the pixels are not evaluated correctly. In this section, we have used the classification error matrix to analyze the segmentation error of image areas. The analysis results of the used database images will be shown in this section. Figure 6 shows the error matrix analysis on a sample image from the Indian Pines database. Figure (6a) shows the ground truth for the extracted regions. Figure (6b) shows the map of the separated areas. Figure (6c) in this image shows the wrong detection areas in color. Figure (6 d) in this image shows the reference data guide for each region.



**Fig. 6.** Analysis of a sample image from the Indian Pines database. A) Ground truth in extracted areas. b) Separated regions (c) Image of misdiagnosis regions. (d) Guide to reference data for each region.

#### 4.2. Compare With Other Articles

The proposed method manages the uncertainty in identifying image regions by using the proposed two-stage SegNet structure. In fact, in the images prepared in different fields, the objects may not have clear boundaries, and according to the conditions of the image, they cannot be easily distinguished from each other. Two-stage segmentation in SegNet provides accurate distinction between objects and regions of the image, and as a result, accurate demarcation in region detection is achieved. To achieve an accurate and fair evaluation, the results of the proposed method in this research have been compared with the results reported in several articles and similar research works. In this evaluation, the criterion under consideration is accuracy. The values presented in the results obtained from CNN algorithms, valid patch, Watershed + SVM method, clustering method (SVM), deep automatic encoder method, region expansion method, DBN method, active learning method, recurrent neural network method and Also, the proposed method in this research is shown in Table 1.

### 19. CONCLUSION

In this research, a new SegNet architecture for semantic segmentation of HSI remote sensing images is presented. The proposed method uses a two-stage hierarchical configuration of FCNNs, especially SegNet-type networks, to achieve superior segmentation accuracy. As the simulation results show, the overall accuracy and F1 score results of the proposed scheme are significantly higher than the original structure of FCNN and some other competitors in the category of deep networks. This advantage also includes the detection of small classes and all are obtained without any post-processing. Also, the ROC evaluation shows that the results of the proposed method are statistically reliable.

**Table 1.** Comparing the accuracy of the results obtained from the proposed method with several other algorithms for the classification of the University of Pavia database.

Reference	Method	Accuracy obtained (average of cases)
[23]	Watershed + SVM method	83/10
[24]	Validated segmentation method	62/08
[25]	Clustering method (SVM)	73/20
[26]	Deep Auto encoder method	97/98
[27]	Region expansion method	91/50
[28]	DBN method	88/00
[29]	Active Learning method	73/22
[30]	Recurrent Neural Networks method	86/33
--	The proposed method in this research	92/44

As can be seen in Table 1, the suggested approach does not show the best output compared to other existing methods. But the accuracy obtained shows that this method is widely accepted and has a performance close to strong methods such as deep networks and recurrent networks.

## REFERENCES

- [1] Chakraborty, R., R. Sushil, and M.L. Garg, *Hyper-spectral image segmentation using an improved PSO aided with multilevel fuzzy entropy*. Multimedia Tools and Applications, 2019. **78**(23): p. 34027-34063.
- [2] Emadi, M., Z. Jafarian Dehkordi, and M. Iranpour Mobarakeh, *Improving the Accuracy of Brain Tumor Identification in Magnetic Resonance imaging using Super-pixel and Fast Primal Dual Algorithm*. International Journal of Engineering, 2023. **36**(3): p. 505-512.
- [3] Sathiyamoorthi, V., et al., *An effective model for predicting agricultural crop yield on remote sensing hyper-spectral images using adaptive logistic regression classifier*. Concurrency and Computation: Practice and Experience: p. e7242.
- [4] Abbasi, A., et al., *A meta-analysis of factors related to fertility attitudes, desires, and childbearing intentions in Iranian studies*. Interdisciplinary Studies in Humanities, 2022. **14**(4): p. 63-92.
- [5] Harouni, M., M. Karimi, and S. Rafieipour, *Precise segmentation techniques in various medical images*. Artificial Intelligence and Internet of Things: Applications in Smart Healthcare, 2021. **117**.
- [6] Karimi, E., A. Ebrahimi, and M.R. Tavakoli, *How optimal PMU placement can mitigate cascading outages blackouts?* International Transactions on Electrical Energy Systems, 2019. **29**(6): p. e12015.
- [7] Karimi, M., et al., *Automatic lung infection segmentation of covid-19 in CT scan images*, in *Intelligent Computing Applications for COVID-19*. 2021, CRC Press. p. 235-253.
- [8] Mahmudi, F., M. Soleimani, and M. Naderi, *Some Properties of the Maximal Graph of a Commutative Ring*. Southeast Asian Bulletin of Mathematics, 2019. **43**(4).
- [9] Srinivas, B. and J.R. Prasad, *Enhanced Segmentation Algorithm for Hyper-spectral Imaging (HSI)*.
- [10] Harouni, M., et al., *Health monitoring methods in heart diseases based on data mining approach: A directional review*, in *Prognostic models in healthcare: Ai and statistical approaches*. 2022, Springer. p. 115-159.
- [11] Karimi, M., et al., *Improving monitoring and controlling parameters for alzheimer's patients based on iomt*, in *Prognostic models in healthcare: Ai and statistical approaches*. 2022, Springer. p. 213-237.
- [12] Jia, J., et al., *Review on active and passive remote sensing techniques for road extraction*. Remote Sensing, 2021. **13**(21): p. 4235.
- [13] Karimi, M., M. Harouni, and S. Rafieipour, *Automated medical image analysis in digital mammography*, in *Artificial intelligence and internet of things*. 2021, CRC Press. p. 85-116.
- [14] Navabifar, F. and M. Emadi, *A Fusion Approach Based on HOG and Adaboost Algorithm for Face Detection under Low-Resolution Images*. INTERNATIONAL ARAB JOURNAL OF INFORMATION TECHNOLOGY, 2022. **19**(5): p. 728-735.
- [15] Rehman, A., et al., *Microscopic retinal blood vessels detection and segmentation using support vector machine and K-nearest neighbors*. Microscopy research and technique, 2022. **85**(5): p. 1899-1914.
- [16] Soleimani, M., M.H. Naderi, and A.R. Ashrafi, *TENSOR PRODUCT OF THE POWER GRAPHS OF SOME FINITE RINGS*. Facta Universitatis, Series: Mathematics and Informatics, 2019: p. 101-122.
- [17] Moshayedi, A.J., et al., *E-Nose design and structures from statistical analysis to application in robotic: a compressive review*. EAI Endorsed Transactions on AI and Robotics, 2023. **2**(1): p. e1-e1.
- [18] Shafique, A., et al., *Deep learning-based change detection in remote sensing images: a review*. Remote Sensing, 2022. **14**(4): p. 871.
- [19] *High-level hyperspectral image classification based on spectro-spatial dimensionality reduction*. Spatial Statistics, 2016: p. 103-117.
- [20] *Hyperspectral image classification with SVM*. Wireless Communications and Networking 2019: p. 1-9.
- [21] *Texture Pattern Separation for Hyperspectral*. Selected Topics in Applied Earth Observations and Remote Sensing 2019: p. 3602 - 3614.
- [22] *KNN-Based Representation of Superpixels for Hyperspectral Image Classification*. SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, 2018: p. 1-16.
- [23] *Local Binary Pattern-Based Hyperspectral Image Classification With Superpixel Guidance*. TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, 2018. **56**(2): p. 749-759.
- [24] Huang, W., et al., *Local binary patterns and superpixel-based multiple kernels for hyperspectral image classification*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020. **13**: p. 4550-4563.
- [25] Jia, S., et al., *Local binary pattern-based hyperspectral image classification with superpixel guidance*. IEEE Transactions on Geoscience and Remote Sensing, 2017. **56**(2): p. 749-759.
- [26] Li, W., et al., *Feature extraction for hyperspectral images using local contain profile*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019. **12**(12): p. 5035-5046.
- [27] Sun, C. *Analyzing Multispectral Satellite Imagery of South American Wildfires Using Deep Learning*. in *2022 International Conference on Applied Artificial Intelligence (ICAPAI)*. 2022. IEEE.
- [28] Boonpook, W., et al., *Deep Learning Semantic Segmentation for Land Use and Land Cover Types Using Landsat 8 Imagery*. ISPRS International Journal of Geo-Information, 2023. **12**(1): p. 14.

- [29] Kaul, A., *A Fundamental Review on Hyperspectral Segmentation Algorithms*, in *Applications of Networks, Sensors and Autonomous Systems Analytics*. 2022, Springer. p. 165-185.
- [30] Balafar, M.A., et al., *Review of brain MRI image segmentation methods*. *Artificial Intelligence Review*, 2010. **33**(3): p. 261-274.
- [31] Soleimani, M., F. Mahmudi, and M. Naderi, *Some results on the maximal graph of commutative rings*. *Advanced Studies: Euro-Tbilisi Mathematical Journal*, 2023. **16**(suppl1): p. 21-26.
- [32] Chen, Y., et al., *Deep feature extraction and classification of hyperspectral images based on convolutional neural networks*. *IEEE Transactions on Geoscience and Remote Sensing*, 2016. **54**(10): p. 6232-6251.
- [33] Tarabalka, Y., J. Chanussot, and J.A. Benediktsson, *Segmentation and classification of hyperspectral images using watershed transformation*. *Pattern Recognition*, 2010. **43**(7): p. 2367-2379.
- [34] Nalepa, J., M. Myller, and M. Kawulok, *Validating hyperspectral image segmentation*. *IEEE Geoscience and Remote Sensing Letters*, 2019. **16**(8): p. 1264-1268.
- [35] Bilgin, G., S. Erturk, and T. Yildirim, *Segmentation of hyperspectral images via subtractive clustering and cluster validation using one-class support vector machines*. *IEEE Transactions on Geoscience and Remote sensing*, 2011. **49**(8): p. 2936-2944.
- [36] Chen, Y., et al., *Deep learning-based classification of hyperspectral data*. *IEEE Journal of Selected topics in applied earth observations and remote sensing*, 2014. **7**(6): p. 2094-2107.
- [37] Li, F., et al., *ST-IRGS: A region-based self-training algorithm applied to hyperspectral image classification and segmentation*. *IEEE Transactions on Geoscience and Remote Sensing*, 2017. **56**(1): p. 3-16.
- [38] Li, T., J. Zhang, and Y. Zhang. *Classification of hyperspectral image based on deep belief networks*. in *2014 IEEE international conference on image processing (ICIP)*. 2014. IEEE.
- [39] Huck, J.J., et al., *Centaur VGI: A Hybrid Human-Machine Approach to Address Global Inequalities in Map Coverage*. *Annals of the American Association of Geographers*, 2021. **111**(1): p. 231-251.
- [40] Shah Heydari, S., *Large Area Land Cover Mapping Using Deep Neural Networks and Landsat Time-Series Observations*. 2021.



# Neural Network Design for Energy Estimation in Surge Arresters

Zohreh Dorrani<sup>1</sup> , Hojat Jannat Abadi<sup>2</sup>

1, 2- Department of Electrical Engineering, Payame Noor University (PNU), Tehran, Iran.

Email: dornai.z@pnu.ac.ir (Corresponding author)

## ABSTRACT:

In power systems, the transmission and distribution networks of electrical energy rely heavily on the performance of various equipment. Any malfunction within these systems can lead to network interruptions, short circuits, and power failures. Arresters are critical devices used to limit transient overvoltages caused by lightning strikes and switching events in transmission and distribution networks. These arresters protect equipment from transient overvoltages while ensuring that they do not react to temporary overloads. Their effectiveness is influenced by environmental conditions, such as humidity and pollution. This research aims to analyze the factors affecting voltage and energy absorption during lightning strikes on power systems. Additionally, we focus on designing an artificial neural network (ANN) to estimate the energy absorbed by the arrester, minimizing the error of this neural network. The results demonstrate that the ANN can effectively estimate the power of the arrester within the power system, providing a valuable tool for enhancing system reliability and performance. This study contributes to the understanding of arrester behavior under transient conditions and offers a novel approach to estimating their energy absorption capabilities using advanced computational techniques.

**KEYWORDS:** Arresters, Lightning, Neural Network, Power Systems.

## 20. INTRODUCTION

The lightning strike is a natural phenomenon that has long been the source of many human and financial injuries. This phenomenon is the main reason to cause line tripping and service interruption, the many numbers of accidents in the power system are caused by lightning which has brought a great loss to the national economy [1]. Lightning is an enormous flash resultant from the increase of millions of volts between clouds or between a cloud and the earth [2].

To prevent the destructive effects of lightning, a suitable earth system was proposed [3]. The earth system is to provide the electric earth with low-resister that can protect the equipment against electric shock. The effect of grounding resistance connected to surge arresters [4] was presented and the result shows that the grounding resistance of Surge Arresters can be increased to some extent without decreasing the lightning protection level [5]. Today, the surge arrester is often used to protect the equipment of the power system against transient overvoltages.

Lightning rod arrester [6] is used to keep the tower and present another tradition for lightning current. This arrester has reduced the potential of tower overhead and the performance of a 500 kV lightning rod arrester is tested, and employed in transmission arrangement. When it is linked to the tower, the top possible of the tower can be limited. Understanding the nature of overvoltages and mobile waves makes it possible for manufacturers to design the appropriate level of insulation to protect the power systems [7]. The back flashover, direct lightning hit to a phase conductor, and lightning-induced voltage are the category of lightning overvoltages. A statistical technique was used to investigating the energy absorption of each surge arrester to take into account the lightning parameter randomness [8]. A probabilistic evaluation of the energy absorption capability of transmission line surge arresters (SAs), based on the Monte-Carlo method was presented.

We propose investigating the factors affecting the voltage and energy absorbed by the arrester when the lightning strikes the system power and the design of an artificial neural network [9] to estimate the energy of the arrester. So that the error neural network [10] can be neglected and it is possible to use this neural network to estimate the power of the arrester in the power system.

The energy source estimation using the neural network [11] has not been done before. With this method, the accuracy level. To study the characteristics of power line arrays in high voltage substations and power lines, the EMTP Works

software [12] is a very useful tool for simulating the power system. So, using EMTP Works software, we first identified the effective factors on the amount of voltage and energy depleted on the arrester when the lightning strikes into the power system. Then, after identifying the factors affecting the voltage and energy of the arrester, with the help of the MATLAB software, we designed an artificial neural network [13]. The neural network can be neglected and it is possible to estimate the power of the arrester in the power system. The arrester must have sufficient absorption capacity to withstand the thermal shock caused by the shock absorber discharge. A choice of energy capacity for a surge arrester depends on many factors, including practical experience, statistics on network connections, storm statistics with lightning strikes and information about the line drainage class. Choosing the right amount for the energy of an arrester is very difficult, which is possible using the neural network [14] in this paper. This network has an amount of non-critical error to predict the energy of the archer so this network Nervous system as a power surge arrester in the power system.

**21. MATERIALS AND METHODS**

The ZnO arresters [15] have important dynamic and frequency characteristics for lightning waves and other waves with a rapid wavefront. In the simulation, the arrester is simulated with a nonlinear resistance. The lightning waves have a fast front slope therefore; the dynamic effects are provided for the ZnO arresters. Fig. 1 shows the proposed model for The ZnO arrester. For waves with a slow front, the RL filter shows the small impedance. The value of the impedance of the RL filter is very important in waves with high-speed wavefront [16]. The practical arrester data be given by:

$$L_0 = 0.33\mu H, R_0 = 170\Omega, L_1 = 32\mu H, R_1 = 105\Omega, C = 0.031nH$$

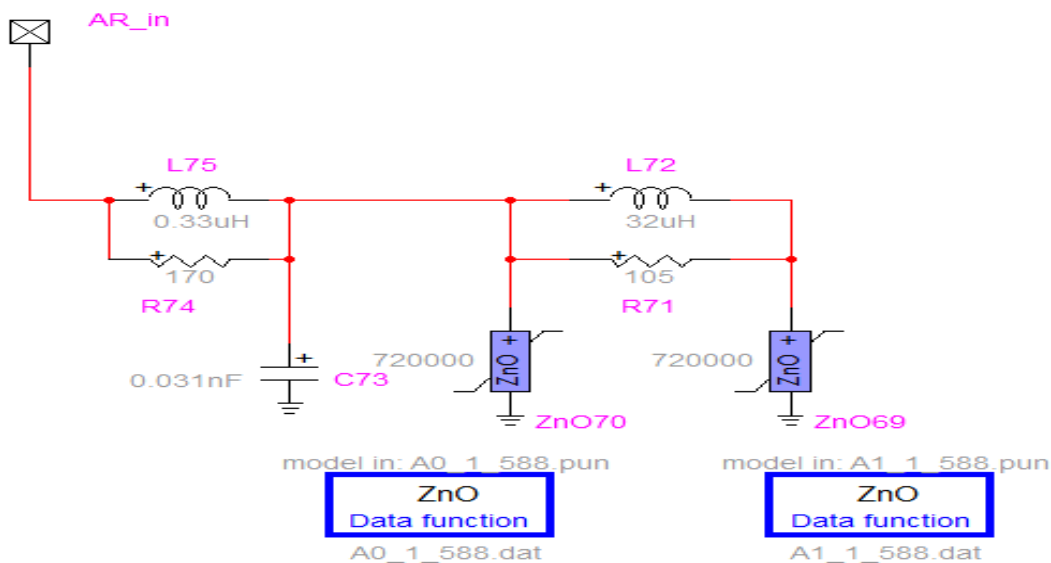


Fig. 1. Frequency model of ZnO arrester.

$$R_0 = \frac{100d}{n} \Omega, L_0 = \frac{0.2d}{n} \mu H, R_1 = \frac{65d}{n} \Omega$$

$$[20] C = \frac{100n}{d} PF, L_1 = \frac{15d}{n} \mu H$$

Where d is the length of the arrester in meters, n is the number of parallel columns consisting of disks ZnO, L<sub>0</sub> is magnetic inductance due to adjacent fields of arrester or ring inductance including transformer and arrester, R<sub>0</sub> is to

stabilize integral calculations and  $c$  is a capacitor of the arrester. To calculate the energy of the archer, the power of the arrester is calculated then, the energy of the archer can be achieved following equation:

$$E(t) = \int_0^t P(\tau) d\tau \quad (2)$$

Where  $E(t)$  is the energy and  $P(\tau)$  is the power. The archer's energy mainly depends on three factors, the current intensity of the lightning wave, the sequence duration time of the lightning wave behind, and the resistance of the tower foot. The current intensity of the lightning wave and the sequence duration time of the lightning wave behind have a random nature and the resistance of the tower foot has a selective nature. Therefore, by changing the resistance of the tower foot, the stress of the energy entered by the arrester can be changed.

## 22. RESULTS

When lightning happens, it may collision to phase conductor or wire guard. The simulation operation was performed for the lightning collision to once to the wire guard and once to the phase conductor.

To investigate and compare the affecting factors on the voltage and energy of the arrester, before the simulation, a base state was required. The base state should not be too small or too large. If these parameters are small, the factors affecting the voltage and the energy of the arrester are not possible. If the parameters of the base state are large, we will give the incorrect and unrealistic values of the voltage and energy values of the arrester. These parameters can be achieved with a test and error model. We consider the base state as Table 1.

**Table 1.** The base state parameters for investigating the voltage and energy of the arrester.

$I_{\text{surge (base)}} \text{ (KA)}$	$t_f \text{ (base)} \text{ (}\mu\text{s)}$	$t_h \text{ (base)} \text{ (}\mu\text{s)}$	Span Length <sub>(base)</sub> (m)	$R_{\text{foot (base)}} \text{ (}\Omega\text{)}$
100	8	200	450	200

The simulation results were obtained for measuring the voltage and energy when a lightning strike hit the wire guard and phase conductor. It is estimated that by gathering this information, the amount of network load can be measured to appropriately design and utilize arresters. All energy values in the tables are expressed in joules, and all voltage values are presented in kilovolts.

**Table 2.** Energy stress of the arrester with change in the intensity of the lightning current.

I(KA)	10	20	35	50	80	100	120	140
E (Guard wire)	2	3	8	642	10915	22253	36845	5401
E (Phase inductor)	314819	1279274	2991618	5079131	9792491	13196658	16811545	20610291

As can be seen from the table above, if the other parameters remain constant, E (Guard wire) initially increases and then decreases. The sharpness of E (Phase inductor) increases.

**Table 3.** Voltage stress of the arrester with change of the intensity in the lightning current.

I(KA)	10	20	35	50	80	100	120	140
V (Guard wire)	3	11	33	118	281	386	490	582
V(Phase inductor)	834	907	962	1015	1126	1182	1231	1273

Table 3 shows that if the lightning current is variable, the sharpness of E (Guard wire) and E (Phase inductor) increases.

**Table 4.** Energy stress of the arrester is presented with changes in the duration time of the lightning wave.

$t_h(\mu s)$	20	50	100	150	200	250
E (Guard wire)	3	1999	11390	18723	22253	23763
E (Phase inductor)	928714	3615554	7505276	10521206	13196658	15593095

In this table, with increasing time behind sequence duration time of the lightning wave, (Phase inductor) increases.

**Table 5.** Voltage stress of the arrester with change behind sequence duration time of the lightning wave.

$t_h(\mu s)$	20	50	100	150	200	250
V (Guard wire)	315	362	379	384	386	387
V(Phase inductor)	1173	1181	1193	1205	1219	1226

Table 5 demonstrates that as the duration time of the lightning wave increases, both V (Guard wire) and V (Phase inductor) increase.

**Table 6.** Energy stress of the arrester with a change of resistance of the tower foot.

$R_{foot}(\Omega)$	50	100	200	300	400	500
E (Guard wire)	1403	22253	74878	122333	159079	186089
E (Phase inductor)	13206049	13196658	13192783	13188450	13185333	13183057

**Table 7.** Voltage energy stress of the arrester with change of resistance of the tower foot.

$R_{foot}(\Omega)$	50	100	200	300	400	500
V (Guard wire)	172	386	662	712	742	762
V(Phase inductor)	1183	1174	1150	1139	1128	1114

The effect of tower footing resistance on the voltage and energy of the lightning arrester is illustrated in Tables 6 and 7. It is summarized that when a lightning strike hits a phase inductor, the arrester's energy and voltage decrease as the tower footing resistance increases. An inverse relationship exists between tower footing resistance and the voltage and energy of the lightning arrester. Conversely, when a lightning strike impacts the guard wire, an increase in tower footing resistance leads to a rise in both the energy stress and voltage of the arrester. In this scenario, a direct relationship is observed.

According to the results, the main factors controlling energy absorption have been identified. Currently, neural networks are employed for various pattern-recognition tasks, including line recognition, speech recognition, and image processing. They are also utilized for classification issues such as text or image categorization. An artificial neural network has been designed to estimate the energy absorbed by the arrester, focusing on minimizing the error of this neural network. It is noteworthy that the neural network can effectively estimate the energy of the arrester within power systems.

Although deep learning is a newer approach today, the use of simple neural networks for estimating the energy of surge arresters offers significant advantages. One of these advantages is high prediction accuracy. Neural networks are capable of learning complex patterns and relationships within data, leading to more precise predictions of energy absorption during lightning strikes. Additionally, these networks can be trained on historical data, allowing them to adapt to various environmental conditions and changes in lightning characteristics.

Moreover, efficiency in data processing is another benefit of using neural networks. These networks can process large volumes of data quickly and effectively, enabling real-time analysis and decision-making. Furthermore, by automating the estimation process, the likelihood of human error in calculations and assessments related to surge arresters is reduced. Ultimately, the accuracy in estimating the energy absorbed by surge arresters contributes to enhancing the reliability and stability of electrical systems.

The use of simple neural networks has significant advantages over deep learning methods. One of these advantages is the simplicity and speed of training. Simple neural networks typically have fewer structures and therefore require less computational resources. This results in reduced training time and a faster model development process, especially in projects that require quick implementation [17].

Additionally, the reduced risk of overfitting is another benefit of using simple neural networks. With fewer parameters in these types of networks, the likelihood that the model becomes overly dependent on the training data and fails to generalize well to new data is lower. This characteristic is particularly important in applications where training data is limited or insufficient [18]. Although deep learning and artificial intelligence are suitable methods, they can be utilized in future research.

The number of neurons in the hidden layer was determined using the test and error method. After testing, 10 neurons were selected as the optimal number that produced the best convergence between the generated results and the training data. The neural network consists of three layers: an input layer containing 3 neurons, a hidden layer containing 10 neurons, and an output layer containing 1 neuron (Fig. 2).

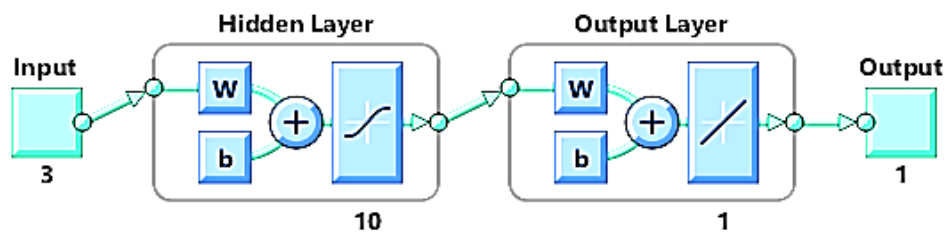


Fig. 2. MLP neural network designed.

Selecting input-output data is very important for network education. This information should include various conditions and conditions that may occur in the actual system so that the neural network [19] experiences different conditions and is resistant to various inputs. The current intensity of the lightning wave, the sequence duration time of the lightning wave behind, and the resistance of the tower foot are inputs and the energy of the arrester is the output factor of the neural network [20] model. (Fig. 3)

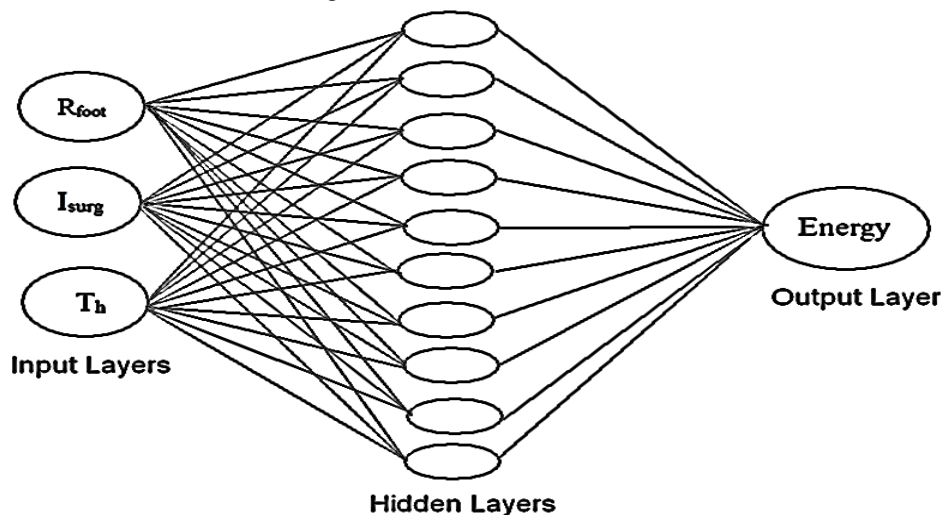


Fig. 3. The general scheme of the MLP neural network.

The amount of the lightning current between 4-140KA, the sequence duration time of the lightning wave behind between 20-250 $\mu$ s, and the R foot value increased by 9 steps and at each step of 50 ohms, ranging from 100 ohms to 500 ohms.

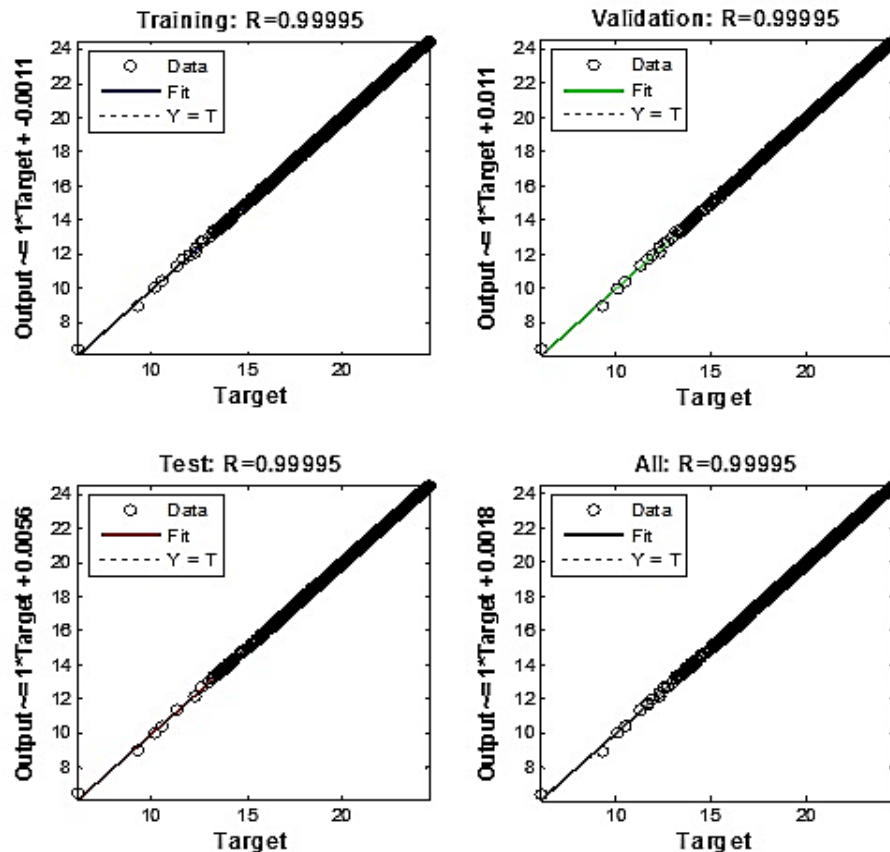


Fig. 4. Regression status of learning, testing, verification, and data allowed data when hit lightning phase conductor.

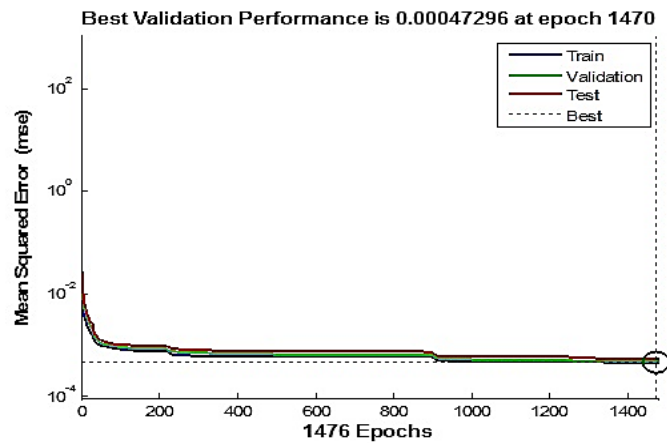
The energy values of the arrester were obtained in two scenarios: when lightning struck the phase conductor and when it impacted the wire guards. For each scenario (lightning strike on the phase conductor and wire guard), a total of 29,187 input/output samples were considered for network training. To enhance training efficiency and accommodate the energy array sizes, all input-output data were normalized using the non-linear  $\log_2 X$  function within the interval (2.24).

After designing the network, simulated results were obtained through the inverse method of normalization. It was noted that normalizing the results caused them to differ from the original values. To retrieve the original results, a reverse normalization process was necessary.

Once the neural estimator was developed and subjected to various inputs while determining appropriate network weights, it could be utilized to estimate the energy absorbed by the arrester in the power grid. In the designed neural network, which aims to predict the energy of the arrester during collisions, the regression status of training, testing, and verification data is illustrated for lightning strikes on the phase conductor in Fig. 4. This graph displays the output in relation to the target. The vertical axis represents the optimal measured value and the maximum coefficient ( $\mu$ ), which indicates the degree of difference that should be reflected in the diagram. The target variable corresponds to impacts with either a guard wire or phase conductor, varying from 0 to 25.

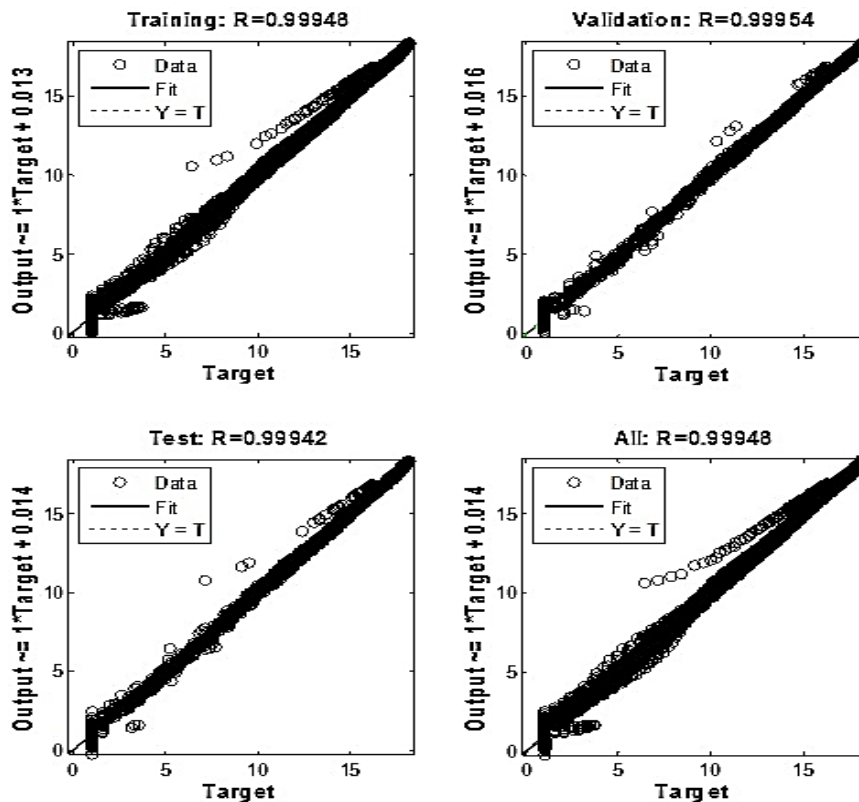
Ideally, when the calculated outputs are almost in line with the target outputs, the graph is in a straight line with a gradient of 45 degrees and can be used to accurately the results. The mean square error (MSE) [21] in the training, test, and confirmation data was illustrated for each repetition in Fig. 4. The vertical axis gives the average output error of the applied 3 data. This figure is the best condition where the error rate is close to real.





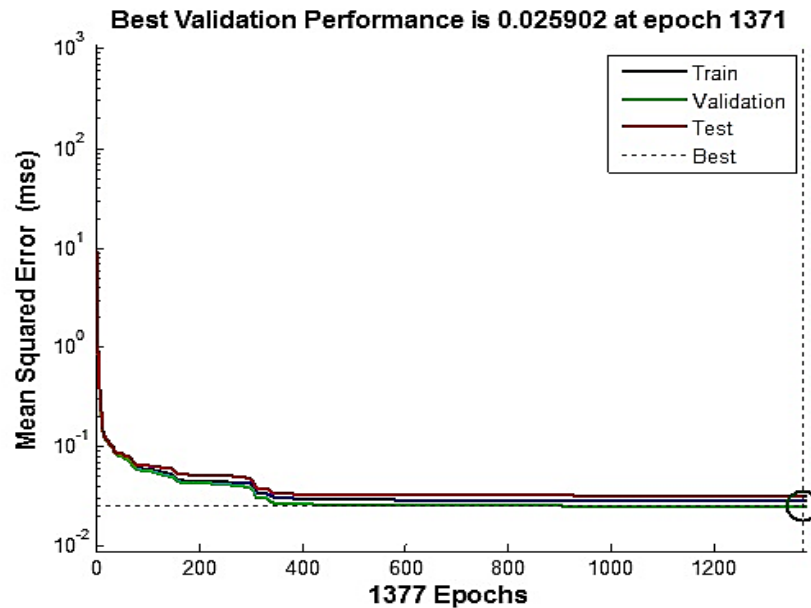
**Fig. 5.** MSE status in educational, testing, and verification data for each repetition when attacking the lightning to the phase conductor.

In Fig. 5, the average errors of confirmed data, test data, and historical data (real) are represented by the green, red, and blue colors, respectively. The error rate is calculated for each iteration, and the optimal mode, where the error rate closely aligns with reality, is indicated by a circle in the diagram. It is observed that after 1,476 iterations, no further changes occur, and the error stabilizes at that point. This indicates that the network's error is unlikely to decrease further after 1,476 iterations. Therefore, lightning strike data can be effectively integrated into the power network with a good fit.



**Fig. 6.** Regression status of learning, testing, verification and data allowed data when hit lightning to the wire guard.

Figures 6, and 7 show data regression and the mean squared error of output, respectively when the lightning strikes the guard wire.



**Fig. 7.** MSE status in educational, testing, and verification data for each repetition when attacking the lightning to the wire guard.

In this case, the network error after 1371 times, there is no longer the probability of it, and so it is possible to easily load the loaded lightning with a good ratio to the power grid.

### 23. CONCLUSIONS

The results of this study highlight the critical role of surge arresters in limiting transient overvoltages in power systems. By utilizing EMTP Works software, the transient states within the network were accurately simulated, allowing for a detailed analysis of the voltages and energies absorbed by the arresters during lightning strikes. It was determined that the energy and voltage experienced by the arresters are influenced by three key factors: the intensity of the lightning wave, the duration of the lightning wave, and the resistance at the tower's base.

Through this research, significant insights into the parameters affecting arrester performance were gained. The implementation of a powerful neural network provided a robust tool for estimating energy surges, yielding a low error rate in predictions. This advancement demonstrates that neural networks can effectively model complex behaviors in power systems, ultimately enhancing the reliability and protection offered by surge arresters against transient overvoltages.

### REFERENCES

- [1] N. Ravichandran, D. Proto, and A. Andreotti, "Surge arrester optimal placement in distribution networks: A decision theory-based approach," *Electric Power Systems Research*, vol. 234, p. 110744, 2024.
- [2] L. Cai *et al.*, "Electromagnetic fields of lightning return stroke to wind turbines with discontinuous impedance model," *Electric Power Systems Research*, vol. 233, p. 110515, 2024.
- [3] M. Boukhouna, B. Nekhoul, and B. Khelifi, "Time domain modeling of lightning transients in grounding systems considering frequency dependence and soil ionization," *Electric Power Systems Research*, vol. 234, p. 110542, 2024.
- [4] B. Ranjbar, A. Darvishi, R. Dashti, and H. R. Shaker, "A survey of diagnostic and condition monitoring of metal oxide surge arrester in the power distribution network," *Energies*, vol. 15, no. 21, p. 8091, 2022.
- [5] M. Khodsuz and V. Mashayekhi, "Grounding system impedance influence on the surge arrester frequency-dependent model parameters using PSO-GWO algorithm," *COMPEL-The international*

- journal for computation and mathematics in electrical and electronic engineering*, vol. 42, no. 6, pp. 1456-1476, 2023.
- [6] M. Zainuddin and L. Bima, "**Jarak Penempatan Lightning Arrester sebagai Pelindung Transformator terhadap Tegangan Lebih pada Gardu Induk 150 Kv Harapan Baru,**" *Mutiara: Jurnal Ilmiah Multidisiplin Indonesia*, vol. 1, no. 2, pp. 164-185, 2023.
- [7] S. Xu, H. Tu, and Y. Xia, "Resilience enhancement of renewable cyber-physical power system against malware attacks," *Reliability Engineering & System Safety*, vol. 229, p. 108830, 2023.
- [8] A. H. K. Asadi, M. Eskandari, and H. Delavari, "**Accurate Surge Arrester Modeling for Optimal Risk-Aware Lightning Protection Utilizing a Hybrid Monte Carlo-Particle Swarm Optimization Algorithm,**" *Technologies*, vol. 12, no. 6, p. 88, 2024.
- [9] H. Abduljabar Salim Ahmed and R. Asgarnezhad, "**Improving students' performance prediction using LSTM and neural network,**" *Majlesi Journal of Telecommunication Devices*, vol. 12, no. 3, pp. 121-127, 2023.
- [10] S. M. M. Ziaei, P. Etezadifar, Y. Nouruzi, and N. Zarei, "**Distinction of Target and Chaff Signals by Suggesting the Optimal Waveform in Cognitive Radar using Artificial Neural Network,**" *Majlesi Journal of Telecommunication Devices*, vol. 12, no. 2, pp. 69-77, 2023.
- [11] A. Arshaghi and M. Norouzi, "A Survey on Face Recognition Based on Deep Neural Networks," *Majlesi Journal of Telecommunication Devices*, 2023.
- [12] E. Karami, E. Hajipour, M. Vakilian, and K. Rouzbehi, "Analysis of Frequency-Dependent Network Equivalents in Dynamic Harmonic Domain," *Electric Power Systems Research*, vol. 193, p. 107037, 2021.
- [13] Z. Dorrani, "Road Detection with Deep Learning in Satellite Images," *Majlesi Journal of Telecommunication Devices*, vol. 12, no. 1, pp. 43-47, 2023.
- [14] A. A. Abed and M. Emadi, "**Detection and Segmentation of Breast Cancer Using Auto Encoder Deep Neural Networks,**" *Majlesi Journal of Telecommunication Devices*, vol. 12, no. 4, pp. 209-217, 2023.
- [15] R. Rohini and C. Pugazhendhi Sugumaran, "**Enhancement of electro-thermal characteristics of micro/nano ZnO based surge arrester,**" *Journal of Electrical Engineering & Technology*, vol. 16, pp. 469-481, 2021.
- [16] V. Hinrichsen, "**Metal-oxide surge arresters in high-voltage power systems,**" *Fundamentals. Siemens AG, Erlangen, Germany*, 2012.
- [17] Z. Cui, L. Wang, Q. Li, and K. Wang, "**A comprehensive review on the state of charge estimation for lithium-ion battery based on neural network,**" *International Journal of Energy Research*, vol. 46, no. 5, pp. 5423-5440, 2022.
- [18] R. Y. Choi, A. S. Coyner, J. Kalpathy-Cramer, M. F. Chiang, and J. P. Campbell, "**Introduction to machine learning, neural networks, and deep learning,**" *Translational vision science & technology*, vol. 9, no. 2, pp. 14-14, 2020.
- [19] Z. Dorrani, H. Farsi, and S. Mohamadzadeh, "**Deep Learning in Vehicle Detection Using ResUNet-a Architecture,**" *Jordan Journal of Electrical Engineering. All rights reserved-Volume*, vol. 8, no. 2, p. 166, 2022.
- [20] Z. Dorrani, H. Farsi, and S. Mohammadzadeh, "**Edge Detection and Identification using Deep Learning to Identify Vehicles,**" *Journal of Information Systems and Telecommunication (JIST)*, vol. 3, no. 39, p. 201, 2022.
- [21] B. Fesl, M. Koller, and W. Utschick, "**On the mean square error optimal estimator in one-bit quantized systems,**" *IEEE Transactions on Signal Processing*, vol. 71, pp. 1968-1980, 2023.

# Authentication Methods in Internet-of-Things Platform: A Comprehensive Review

Gholam Reza Zargar<sup>1</sup>, Hamid Barati<sup>1</sup>, Ali Barati<sup>1</sup>

1- Department of Computer Engineering, Dezful Branch, Islamic Azad University, Dezful, Iran.

Email: gh.zargar@iaud.ac.ir, hamid.barati@iau.ac.ir (Corresponding author), alibarati@iau.ac.ir

## ABSTRACT:

Authentication is a critical aspect of securing Internet-of-Things (IoT) platforms, ensuring only authorized devices and user's access sensitive data and services. One critical aspect of IoT is ensuring secure and anonymous authentication protocols to safeguard sensitive data. This paper presents a comprehensive review of authentication methods specifically tailored for IoT environments. Through an extensive analysis of existing literature, various authentication techniques in IoT environments are explored. The review covers key aspects such as security mechanisms, privacy preservation techniques, scalability, and usability of these protocols. Furthermore, the paper discusses challenges unique to IoT authentication, such as resource constraints, scalability, and resilience against diverse cyber threats. Various authentication protocols and frameworks applicable to IoT ecosystems are analyzed, highlighting their strengths, weaknesses, and suitability for different IoT use cases. Additionally, the review examines recent advancements in authentication technologies like blockchain in the context of IoT security. Insights from this review aim to provide researchers and practitioners with a deeper understanding of IoT authentication methods and inform the development of robust, efficient, and scalable authentication solutions for IoT platforms.

**KEYWORDS:** Authentication, Internet of Things, Security, Privacy Preservation.

## 1. INTRODUCTION

The Internet of Things refers to a network of interconnected physical objects or "things" embedded with sensors, software, and other technologies that enable them to collect and exchange data with other devices and systems over the internet [1]. These objects can range from everyday items such as household appliances, wearable devices, and vehicles to more specialized equipment used in industrial settings, healthcare, agriculture, and beyond [2-3]. The core concept behind the IoT is to create a seamless ecosystem where devices can communicate with each other, share information, and perform tasks autonomously without requiring human intervention [4-5]. This interconnectedness allows for the creation of smart environments where data collected from various sensors can be analyzed and utilized to improve efficiency, enhance decision-making processes, and enable new services and applications [6].

The IoT ecosystem comprises several key components essential for its operation and functionality [7]. Firstly, sensors and actuators are integral devices embedded within physical objects. Sensors collect data from the surrounding environment, while actuators execute actions based on received instructions [7-8]. Connectivity is another vital component, as IoT devices rely on various communication technologies such as Wi-Fi, Bluetooth, cellular networks, and low-power wide-area networks (LPWAN) to transmit data to other devices or centralized systems. Data processing and analytics play a crucial role in deriving insights from the collected data [9]. This involves real-time processing and analysis using cloud computing platforms or edge computing devices to extract meaningful information for decision-making. Applications and services form the next key component, utilizing the data and insights generated by IoT devices to develop a wide range of applications across industries [10]. These applications include smart home automation, remote healthcare monitoring, predictive maintenance in manufacturing, precision agriculture, and smart city initiatives [11-13]. Finally, security and privacy are paramount considerations in the IoT ecosystem. Given the sensitive nature of the transmitted and stored data, robust security measures and privacy protection mechanisms are essential to prevent unauthorized access, data breaches, and misuse of personal information [14]. This ensures the integrity and confidentiality of the data exchanged within the IoT network. Overall, the IoT holds immense potential to revolutionize

various aspects of our lives, offering unprecedented levels of connectivity, efficiency, and convenience [15]. However, it also presents challenges related to interoperability, security, privacy, and ethical considerations that need to be addressed as the IoT continues to evolve and expand [16].

Security in the IoT is critical due to the vast network of connected devices vulnerable to cyber threats [17]. Ensuring IoT security involves several key measures. First, device authentication and authorization protocols are essential to verify the identity and permissions of devices. Secondly, data encryption is crucial to protect the confidentiality and integrity of information transmitted between devices and servers [18]. Additionally, robust security updates and patch management are needed to address vulnerabilities promptly. IoT devices should also implement secure communication protocols like TLS/SSL to safeguard data in transit. Furthermore, network segmentation can limit the impact of breaches by isolating critical systems from potentially compromised devices. Finally, user awareness and privacy protection are vital considerations. Implementing these measures comprehensively is crucial for building trust in IoT systems and safeguarding against evolving cyber threats [19].

Authentication in IoT involves verifying the identity of devices, users, or applications before allowing access to IoT networks or services [20]. Cryptographic techniques such as digital certificates and secure tokens are commonly used for device authentication, ensuring only trusted devices can interact within IoT ecosystems. Biometric authentication, like fingerprint or facial recognition, adds another layer of security for user access to IoT devices [21]. Multi-factor authentication (MFA) is also vital, requiring multiple credentials for verification, enhancing security against unauthorized access. Secure communication protocols like TLS/SSL are integrated into IoT authentication processes to encrypt data during transmission, safeguarding against eavesdropping and tampering. Strong authentication mechanisms are crucial in IoT to mitigate cyber threats, protect privacy, and maintain the integrity of interconnected systems in increasingly complex and dynamic IoT environments [22].

This paper offers a comprehensive examination of security challenges and prerequisites within the IoT framework, employing a layer-oriented strategy. It subsequently conducts a current assessment of diverse authentication methods employed in IoT systems. Employing a multi-faceted classification approach, it evaluates and contrasts existing authentication protocols, elucidating their strengths and weaknesses.

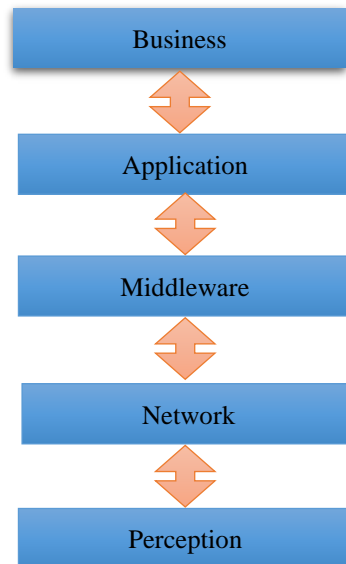
The remainder of the paper is organized as follows: Section 2 describes IoT architecture. Section 3 describes security in IoT. Section 4 presents a taxonomy of IoT authentication schemes. Section 5 reviews previous works on authentication methods in IoT, describing the advantages and disadvantages of each. Section 6 presents the evaluation and comparison of methods. Finally, the conclusion is presented in Section 7.

## 2. IOT ARCHITECTURE

While traditional Internet connects people to a network, IoT has a different approach in which it provides Machine-to-Machine (M2M) and Human-to-Machine (H2M) connectivity, for heterogeneous types of machines in order to support variety of applications [23-24]. Connecting a huge number of heterogeneous machines leads to a massive traffic, hence the need to deal with the storage of big data [25]. Therefore, the TCP/IP architecture, does not suit the needs of IoT regarding various aspects including privacy and security, scalability, reliability, interoperability, and quality of service [26]. Although numerous architectures were proposed for IoT, there is still a need for a reference architecture. The basic architecture model proposed in the literature is a five-layer architecture, as shown in Figure 1. The IoT layered architecture provides a structured framework for designing and implementing IoT systems. It consists of several layers, each serving specific functions to enable the seamless integration of devices, data, and applications in IoT environments [27].

- Perception Layer: This layer comprises sensors, actuators, and other devices that interact with the physical environment to collect data. Sensors gather information such as temperature, humidity, and motion, while actuators perform actions based on received instructions.
- Network Layer: The network layer facilitates communication between devices, allowing them to transmit data to each other or to centralized systems. It includes various communication technologies such as Wi-Fi, Bluetooth, Zigbee, and cellular networks.
- Middleware Layer: The middleware layer provides services for data processing, storage, and management. It includes components such as data brokers, message queues, and protocol converters, which ensure interoperability and scalability in IoT systems.
- Application Layer: The application layer hosts IoT applications and services that utilize the data collected from devices to deliver value-added functionalities. This includes applications for smart home automation, industrial monitoring, healthcare management, and more.
- Business Layer: The business layer encompasses the business logic, rules, and processes that govern IoT operations. It includes components for data analytics, decision-making, and business intelligence, enabling organizations to derive insights and make informed decisions based on IoT data.

By following the IoT Generic Architecture, organizations can design and deploy scalable, interoperable, and secure IoT solutions that effectively harness the power of connected devices to drive innovation and improve efficiency across various industries. Figure 1 shows the 5-layer architecture of the Internet of Things.



**Fig. 1.** Five-layer architecture.

### 3. SECURITY IN IOT

Security issues in the IoT pose significant challenges due to the extensive interconnectivity of devices and diverse deployment scenarios. Common concerns include weak authentication and authorization, leaving devices vulnerable to unauthorized access and control [28]. Inadequate encryption of data transmission exposes sensitive information to interception and compromise. Vulnerabilities in firmware and software, coupled with insecure communication protocols, create opportunities for exploitation by cyber attackers [29]. Physical security risks, such as tampering and theft of devices, further compound these challenges. The absence of standardized security protocols and privacy concerns regarding data collection exacerbate the situation. Addressing these issues requires a comprehensive approach involving robust authentication and encryption mechanisms, regular updates to firmware and software, secure communication protocols, enhanced physical security measures, industry-wide standardization efforts, and prioritization of user privacy and data protection. By adopting these measures, organizations can mitigate the risks associated with IoT deployments and build trust in IoT technologies [30].

#### 3.1. Perception Layer Security Issues and Requirements

The Perception Layer in the IoT ecosystem comprises sensors, actuators, and other devices that interact with the physical environment to collect data. Security issues and requirements in the Perception Layer are critical due to the direct interaction of these devices with the physical world and the sensitive data they collect. Some common security issues and requirements in the Perception Layer include [31-32]:

- **Unauthorized Access:** Without proper authentication mechanisms, malicious actors may gain unauthorized access to sensors or actuators, leading to data manipulation, device tampering, or physical damage.
- **Data Integrity:** Ensuring the integrity of data collected by sensors is essential to prevent tampering or manipulation, which could result in inaccurate or misleading information being processed by IoT systems.
- **Confidentiality:** Protecting the confidentiality of sensor data is crucial, especially in applications where sensitive information such as personal health data or industrial secrets is being collected. Encryption and access control mechanisms can help safeguard sensitive data from unauthorized disclosure.
- **Device Authentication:** Authenticating devices within the Perception Layer is essential to ensure that data is collected from trusted sources. Strong authentication mechanisms, such as digital certificates or secure tokens, can prevent spoofing or impersonation attacks.



- **Physical Security:** Physical security measures are necessary to protect sensors and actuators from physical tampering, theft, or damage. Installing devices in secure locations, implementing tamper-proof enclosures, and monitoring for physical intrusions can help mitigate these risks.
- **Resilience to Environmental Factors:** Sensors deployed in harsh or unpredictable environments may be vulnerable to damage from environmental factors such as extreme temperatures, moisture, or electromagnetic interference. Designing sensors with robust enclosures and protective coatings can enhance their resilience to environmental hazards.

Addressing these security issues requires a multi-faceted approach, including the implementation of strong authentication mechanisms, encryption techniques, physical security measures, and resilience to environmental factors. By addressing these requirements, organizations can ensure the security and reliability of the Perception Layer in IoT deployments.

### 3.2. Network Layer Security Issues and Requirements

The network layer of IoT systems faces several security issues due to the distributed nature of devices and the diverse communication protocols used. Here are some common security issues and requirements for the network layer of IoT [33-34]:

- **Data Confidentiality:** Ensuring that data transmitted over the network is encrypted and only accessible to authorized parties. This prevents eavesdropping and data interception by malicious actors.
- **Data Integrity:** Guaranteeing that data remains unchanged during transmission and reception. This prevents tampering with data in transit, which could lead to unauthorized modifications or disruptions to IoT operations.
- **Authentication and Access Control:** Implementing mechanisms to authenticate devices and users before granting access to IoT networks and resources. This prevents unauthorized devices from joining the network and unauthorized users from accessing sensitive data or controlling devices.
- **Device Identity Management:** Managing and securely storing unique identities for IoT devices to prevent spoofing and impersonation attacks. This ensures that only legitimate devices can communicate with each other and with backend systems.
- **Network Segmentation:** Partitioning IoT networks into separate segments or VLANs to isolate traffic and limit the potential impact of security breaches. This prevents lateral movement by attackers and contains security incidents to specific parts of the network.
- **Firewalls and Intrusion Detection/Prevention Systems (IDS/IPS):** Deploying firewalls and IDS/IPS solutions to monitor and filter network traffic for signs of malicious activity. This helps detect and block unauthorized access attempts, malware, and other network-based threats.
- **Secure Communication Protocols:** Using secure communication protocols such as Transport Layer Security (TLS), Datagram Transport Layer Security (DTLS), and Message Queuing Telemetry Transport (MQTT) with appropriate encryption and authentication mechanisms.
- **Network Traffic Encryption:** Encrypting all network traffic, including data exchanged between IoT devices, gateways, and backend servers. This prevents unauthorized interception and ensures the confidentiality of sensitive information.
- **Secure Remote Management:** Implementing secure methods for remotely managing and updating IoT devices, such as over-the-air (OTA) firmware updates and secure device management protocols. This reduces the risk of vulnerabilities being exploited due to outdated software or misconfigurations.
- **Resilience and Redundancy:** Building resilience into IoT networks through redundancy, failover mechanisms, and disaster recovery plans. This helps mitigate the impact of network outages, DDoS attacks, and other disruptions to IoT operations.

### 3.3. Middleware Layer Security Issues and Requirements

The middleware layer in IoT systems plays a crucial role in processing, aggregating, and managing data collected from the perception layer before it's sent to the application layer. Ensuring security at this layer is essential for protecting sensitive data, maintaining system integrity, and preventing unauthorized access. Here are some common security issues and requirements for the middleware layer in IoT [35-36]:

- **Data Encryption and Secure Data Handling:** Data processed and stored within the middleware layer should be encrypted to protect against unauthorized access and interception. Encryption mechanisms such as Advanced Encryption Standard (AES) should be employed to ensure data confidentiality.

- **Access Control and Authentication:** Implementing robust access control mechanisms to regulate access to middleware components and data. User authentication and authorization mechanisms should be in place to ensure that only authorized individuals or systems can interact with the middleware layer.
- **Secure APIs and Communication Protocols:** Ensuring that APIs and communication protocols used within the middleware layer are secure and resistant to attacks such as injection, tampering, and eavesdropping. Secure protocols like HTTPS and Message Queuing Telemetry Transport (MQTT) with TLS encryption should be used for communication between middleware components.
- **Integrity Checking and Validation:** Implementing mechanisms to verify the integrity of data processed within the middleware layer. Data integrity checks, digital signatures, and message authentication codes (MACs) can be used to ensure that data has not been altered or tampered with during processing.
- **Secure Configuration and Management:** Ensuring that middleware components are configured securely and kept up-to-date with security patches and updates. Secure configuration management practices should be followed to minimize the risk of misconfiguration-related security incidents.
- **Auditing and Logging:** Implementing logging and auditing mechanisms to track and monitor activities within the middleware layer. This helps detect and investigate security incidents, unauthorized access attempts, and other suspicious activities.
- **Secure Integration with External Systems:** When integrating with external systems or third-party services, ensuring that secure integration practices are followed. Secure authentication, data encryption, and secure API endpoints should be used to protect data exchanged between middleware components and external systems.
- **Secure Data Storage:** If the middleware layer stores data temporarily or persistently, ensuring that data is stored securely with appropriate access controls, encryption, and data retention policies. Secure storage mechanisms should be used to protect sensitive data from unauthorized access or disclosure.
- **Resilience and Fault Tolerance:** Building resilience and fault tolerance into the middleware layer to ensure continued operation in the event of system failures, disruptions, or cyberattacks. Redundancy, failover mechanisms, and disaster recovery plans should be in place to minimize downtime and data loss.
- **Security Monitoring and Incident Response:** Implementing security monitoring tools and incident response procedures to detect and respond to security threats and breaches in real-time. Security events should be logged, analyzed, and acted upon promptly to mitigate risks and minimize the impact of security incidents.

### 3.4. Application Layer Security Issues and Requirements

The Application Layer in the IoT ecosystem is responsible for processing and analyzing data collected from sensors and actuators to derive meaningful insights and enable various applications and services. Security issues and requirements at the Application Layer are crucial to safeguard sensitive data and ensure the integrity and availability of IoT systems. Some common security issues and requirements in the Application Layer include [37-38]:

- **Secure Data Storage and Processing:** Ensuring the secure storage and processing of data within IoT applications is essential to prevent unauthorized access, data breaches, and tampering. Encryption techniques and access control mechanisms should be employed to protect sensitive data stored in databases or processed by applications.
- **Authentication and Authorization:** Implementing robust authentication and authorization mechanisms within IoT applications is vital to verify the identity of users and devices and control their access to sensitive resources. Strong authentication methods, such as multi-factor authentication or biometric authentication, can help prevent unauthorized access to IoT applications.
- **Secure Communication Protocols:** Utilizing secure communication protocols, such as HTTPS or MQTT with TLS/SSL, is essential to encrypt data transmission between IoT devices and backend systems. Secure protocols help protect data from interception and eavesdropping by malicious actors and ensure its confidentiality and integrity during transmission.
- **Vulnerability Management:** Regularly updating and patching IoT applications to address known security vulnerabilities is crucial for maintaining their security posture. Vulnerability management processes should be implemented to identify, prioritize, and remediate security flaws in IoT applications in a timely manner.
- **Secure APIs and Interfaces:** Securing APIs and interfaces used by IoT applications to interact with external systems or services is essential to prevent unauthorized access or manipulation of data. Implementing authentication, access control, and encryption mechanisms for APIs and interfaces helps protect sensitive data and prevent security breaches.
- **Data Privacy and Compliance:** Ensuring compliance with data privacy regulations and standards, such as GDPR or HIPAA, is essential for protecting the privacy of user data collected and processed by IoT applications. Implementing

privacy-by-design principles, data anonymization techniques, and data access controls can help ensure compliance with regulatory requirements and protect user privacy.

Addressing these security issues requires a comprehensive approach, including the implementation of secure data storage and processing practices, robust authentication and authorization mechanisms, secure communication protocols, vulnerability management processes, secure APIs and interfaces, and adherence to data privacy and compliance requirements. By addressing these requirements, organizations can enhance the security and trustworthiness of IoT applications and protect sensitive data from unauthorized access or manipulation.

### 3.5. Business Layer Security Issues and Requirements

The Business Layer in the IoT ecosystem is responsible for managing business logic, workflows, and interactions between different components of the IoT system. Security issues and requirements at the Business Layer are crucial to safeguard sensitive business data, ensure the integrity of business processes, and protect against various cyber threats. Some common security issues and requirements in the Business Layer include [39]:

- **Access Control and Authentication:** Implementing robust access control mechanisms and authentication protocols helps control access to business-critical resources and ensures that only authorized users or devices can interact with the business layer. Role-based access control (RBAC), multi-factor authentication, and strong authentication mechanisms help enforce access policies and prevent unauthorized access to sensitive business data and functionalities.
- **Secure Business Logic:** Ensuring the security of business logic and workflows is essential to prevent exploitation by malicious actors seeking to compromise the IoT system. Implementing secure coding practices, input validation, and output encoding techniques helps mitigate the risk of injection attacks, such as SQL injection or code injection, which can lead to unauthorized access or manipulation of business data and processes.
- **Data Privacy and Compliance:** Ensuring compliance with data privacy regulations and standards, such as GDPR or HIPAA, is essential for protecting the privacy of user data collected and processed by IoT applications. Implementing privacy-by-design principles, data anonymization techniques, and data access controls helps ensure compliance with regulatory requirements and protect user privacy.
- **Secure Integration with External Systems:** Integrating IoT systems with external business applications, cloud services, or third-party platforms introduces security risks, such as data breaches or unauthorized access. Implementing secure communication protocols, encryption techniques, and access controls for data exchanged between IoT systems and external systems helps mitigate these risks and ensure the confidentiality and integrity of data transmissions.
- **Business Continuity and Disaster Recovery:** Planning for business continuity and disaster recovery helps mitigate the impact of security incidents or system failures on business operations. Implementing backup and recovery procedures, redundant systems, and failover mechanisms ensures the availability and resilience of critical business functions in the event of disruptions or security breaches.
- **Risk Management and Governance:** Implementing risk management processes and governance frameworks helps identify, assess, and mitigate security risks in the IoT ecosystem. Conducting regular security assessments, vulnerability scans, and compliance audits helps proactively identify and address security vulnerabilities and ensure ongoing compliance with security policies and regulations.

Addressing these security issues requires a comprehensive approach, including the implementation of access control mechanisms, secure business logic, data privacy measures, secure integration practices, business continuity planning, risk management processes, and governance frameworks. By addressing these requirements, organizations can enhance the security and resilience of the Business Layer in IoT deployments, safeguarding sensitive business data and ensuring the integrity of business processes and operations. Table 1 shows the security requirements at each layer of the IoT.

**Table 1.** Security requirements in each layer

Layer	Security Requirements
-------	-----------------------

Perception	Device authentication
	Data integrity
	Privacy protection
	Firmware/software security
	Physical security
Network	Secure communication protocols
	Access control
	Encryption
	Intrusion detection and prevention
	Traffic monitoring and analysis
Middleware	Data confidentiality
	Data integrity
	Authentication and authorization mechanisms
	Protection against middleware vulnerabilities
	Secure data transmission and storage
Application	User data protection
	Secure authentication and authorization
	Encrypted communication channels
	Protection against application-level vulnerabilities
	Secure application programming interfaces (APIs)
Business	Protection of sensitive business data
	Compliance with regulations (e.g., GDPR, HIPAA)
	Access controls and role-based permissions
	Secure financial transactions
	Business continuity planning and risk management

#### 4. A TAXONOMY OF IOT AUTHENTICATION SCHEMES

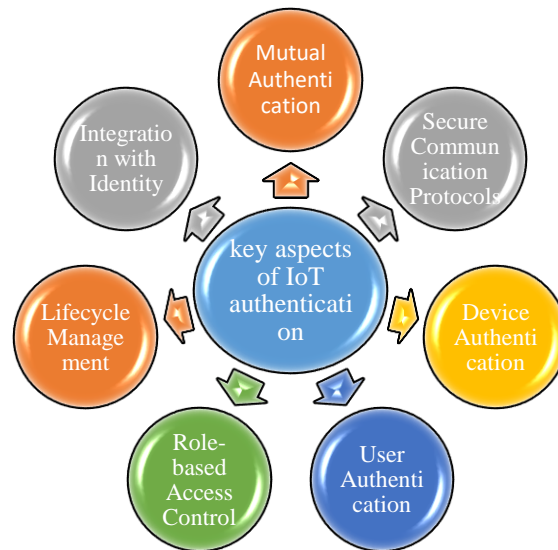
Authentication is the process of verifying the identity of an individual or entity attempting to access a system, network, application, or resource [40]. In other words, it confirms that the person or entity is who they claim to be. Authentication mechanisms typically involve presenting credentials, such as usernames, passwords, biometric data (like fingerprints or facial recognition), security tokens, or cryptographic keys [41]. Authentication is the process of verifying the identity of an individual or entity attempting to access a system, network, application, or resource. Authentication helps ensure that only authorized users gain access to sensitive information or resources, thereby protecting against unauthorized access, data breaches, and security threats [42]. It is a fundamental aspect of cyber security that helps ensure that only authorized users gain access to sensitive information or resources, thereby protecting against unauthorized access, data breaches, and security threats. Authentication is crucial for maintaining the integrity, confidentiality, and availability of information systems and resources. It is used extensively in various domains, including computer systems, networks, websites, mobile devices, and cloud services, to protect against unauthorized access and safeguard sensitive data.

IoT authentication refers to the process of verifying the identity of IoT devices or users interacting with IoT systems, networks, or applications. Given the distributed and heterogeneous nature of IoT environments, authentication becomes crucial for ensuring the security and integrity of IoT deployments. Here are some key aspects of IoT authentication [43]:

- **Device Authentication:** IoT devices need to authenticate themselves to the network or cloud services they interact with. This can involve the use of unique identifiers, such as MAC addresses or cryptographic keys, to establish trust and ensure that only authorized devices can access IoT resources.
- **User Authentication:** In scenarios where users interact with IoT systems through applications or interfaces, user authentication is necessary to verify the identity of individuals accessing the system. This may involve traditional methods like usernames and passwords or more advanced techniques such as biometric authentication.
- **Mutual Authentication:** In some cases, both the IoT device and the server or gateway it communicates with need to authenticate each other to establish a secure connection. This ensures that both parties are legitimate and prevents unauthorized devices or malicious actors from gaining access to sensitive data or control over IoT devices.

- Secure Communication Protocols: IoT authentication often relies on secure communication protocols such as Transport Layer Security (TLS) or DTLS to encrypt data exchanged between devices and servers, protecting against eavesdropping and tampering.
- Role-based Access Control: Access to IoT resources may be restricted based on the roles or permissions assigned to users or devices. Role-based access control (RBAC) mechanisms can enforce authorization policies and ensure that only authorized entities can perform specific actions within the IoT ecosystem.
- Lifecycle Management: Proper authentication in IoT requires managing the lifecycle of devices, including provisioning, registration, deprovisioning, and revocation of credentials. This ensures that only valid and up-to-date devices are allowed to participate in IoT networks.
- Integration with Identity and Access Management (IAM) Systems: IoT authentication mechanisms often need to integrate with existing identity and access management systems to centralize user authentication, enforce security policies, and streamline access control across the entire IoT infrastructure.

The key aspects of IoT authentication are shown in Figure 2.



**Fig. 2.** Key aspects of IoT authentication.

Overall, IoT authentication is essential for establishing trust, preventing unauthorized access, and safeguarding sensitive data in IoT ecosystems. By implementing robust authentication mechanisms, organizations can mitigate security risks and ensure the integrity and reliability of their IoT deployments.

A taxonomy of IoT authentication schemes is a systematic classification framework that categorizes different methods and approaches used for authenticating devices and users within Internet of Things environments. This taxonomy aims to organize the various authentication mechanisms based on their characteristics, functionalities, and deployment models. A taxonomy of IoT authentication schemes categorizes these schemes based on various criteria such as authentication mechanisms, deployment models, security features, and communication protocols. Typically, a taxonomy of IoT authentication schemes includes several categories or dimensions, such as [44]:

- Authentication Mechanisms: Authentication mechanisms in information technology encompass a range of methods to verify the identity of users, devices, or applications. One common approach is password-based authentication, where users provide a secret password to access a system. Biometric authentication uses unique physical characteristics like fingerprints or facial recognition for identity verification, offering a more secure and user-friendly method. Token-based authentication involves the use of physical or virtual tokens, like security keys or one-time password (OTP) tokens, to grant access. Multi-factor authentication combines two or more authentication factors, such as passwords, biometrics, or tokens, to enhance security. Device-based authentication verifies the identity of IoT devices using digital certificates or secure tokens.
- Deployment Models: Authentication deployment models in IT encompass various approaches to how authentication is managed and implemented within systems. Centralized authentication involves a single, central server responsible for verifying user credentials and granting access to resources. This model streamlines management but can be a single point of failure if not properly secured. Distributed authentication spreads the authentication process across

multiple entities or servers, enhancing scalability and resilience. Federated authentication extends authentication across multiple domains or services, allowing users to access resources across different organizations using a single set of credentials. Each deployment model has its advantages and challenges, balancing factors like security, scalability, and ease of use.

- **Security Features:** Security features play a crucial role in ensuring the effectiveness and integrity of authentication schemes in IT environments. Data encryption is fundamental for securely transmitting authentication data over networks, preventing unauthorized access to sensitive information. Mutual authentication enhances security by requiring both parties (user and system) to verify each other's identities before granting access. Key management involves secure storage and distribution of cryptographic keys used for encryption, decryption, and authentication processes, safeguarding against key compromise. Secure communication protocols such as TLS/SSL establish encrypted and authenticated connections between entities, protecting data from interception and manipulation during transmission. Implementing these security features strengthens authentication mechanisms, mitigates risks associated with unauthorized access or data breaches, and fosters trust in the overall security posture of IT systems and services.
- **Communication Protocols:** In the realm of IoT, communication protocols are fundamental for facilitating secure and efficient interactions between IoT devices and authentication servers. Commonly employed protocols include HTTP/HTTPS, which are well-suited for web-based communications with the added security of HTTPS encryption. MQTT is lightweight and efficient, enabling publish-subscribe messaging and supporting secure communication via MQTT over TLS (MQTT-Secure). Constrained Application Protocol (CoAP) is designed specifically for resource-constrained IoT devices, offering low overhead and built-in security features like DTLS for secure data exchange. Advanced Message Queuing Protocol (AMQP) ensures reliable and secure message delivery, making it suitable for industrial IoT applications. Extensible Messaging and Presence Protocol (XMPP) facilitates real-time communication and is ideal for human-to-device interactions in IoT scenarios. DTLS provides security enhancements for UDP-based communication, ensuring confidentiality, integrity, and authentication of data exchanged between IoT devices and authentication servers.

As shown in Figure 3, this taxonomy helps in organizing and understanding the various authentication schemes used in IoT systems, facilitating comparison, selection, and implementation based on specific requirements and constraints. By considering these criteria, a taxonomy of IoT authentication schemes can provide a structured framework for understanding and evaluating the diverse range of authentication methods used in IoT systems.



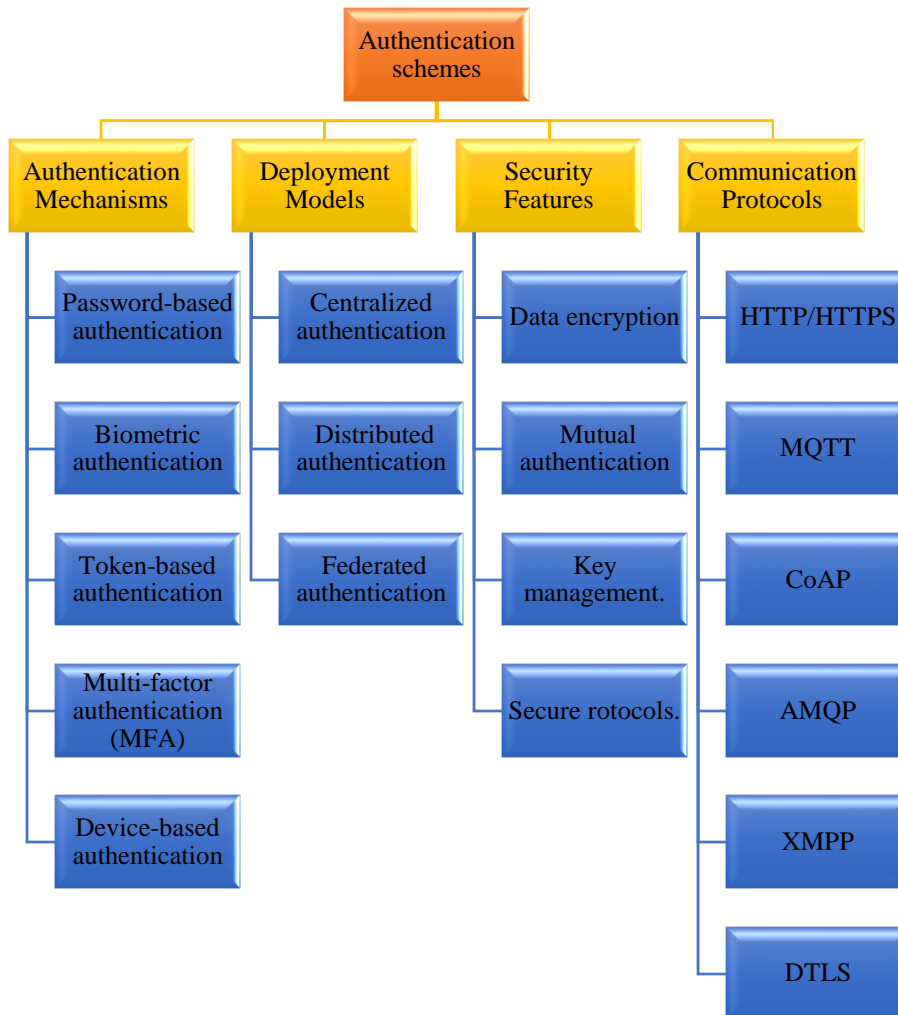


Fig.3. Taxonomy of IoT authentication schemes.

## 5. AUTHENTICATION METHODS

In [45], a lightweight anonymous mutual authentication and key agreement scheme is proposed for two-hop blockchain-based Wireless Body Area Networks (WBANs) to ensure secure message transmission. The scheme enables mutual authentication and key agreement between sensor nodes on patients and various hub nodes across regions, addressing security and lightweight requirements. The protocol is evaluated using the AVISPA tool for security. It consists of Initialization, Registration, and Authentication phases, conducted over secure and public channels. The approach uses XOR operations and a cryptographic hash function for mutual authentication and key agreement. Security analysis confirms robustness. Comparative analysis with related schemes shows improved energy efficiency and security, making it suitable for WBAN implementation.

In [46], two lightweight authentication and key agreement schemes are proposed for IoT device authentication. The first scheme uses ECQV implicit certificates for efficient authentication but lacks public key signature protection. The second scheme enhances security by incorporating Schnorr signatures within CL-PKC, ensuring public key verification. Both schemes leverage identity-based PKC, with the first focusing on implicit authentication and the second integrating signature information into the public user key. These schemes address security concerns like replay attacks and key leakage by minimizing transmitted data and improving communication speed, although the second scheme sacrifices speed for enhanced security.

In [47], a privacy-aware authentication protocol for multi-server CE-IoT systems is proposed, integrating Physical Unclonable Functions (PUFs) and blockchain technology. The protocol encodes real correlations of Challenge-Response Pairs (CRPs) into Mapping Correlations (MCs) using a one-time physical identity and keyed-hash function. Blockchain is employed to store and efficiently synchronize MCs, ensuring secure sharing of physical identities through multi-

receiver encryption. The protocol's security is formally proven using a random oracle model, and its resilience against various attacks is discussed. In the system setup, the Responsible Center (RC) generates public parameters, initializes the Multi-Receiver Encryption (MRE) algorithm, and launches the blockchain system. Authorized verifiers join the blockchain network, and a smart contract is deployed in both the RC and verifiers to manage MCs and CRPs during the registration and authentication phases.

In [48], a secure authentication protocol is presented for a cloud-assisted Telemedicine Information System (TMIS) with access control, integrating blockchain for data integrity. It employs ciphertext-policy attribute-based encryption (CP-ABE) for access control and blockchain to ensure data integrity. The protocol's robustness is demonstrated through informal and Burrows-Adabi-Needham (BAN) logic analyses, along with formal validation using AVISPA, highlighting its security and efficiency advantages over existing methods. Key protocol stages include initialization, registration, key generation, authentication, data upload, treatment, and checkup, defined with corresponding notations in Table 1. To prevent replay attacks, random numbers (secrets) and synchronized timestamps from TMIS entities are used. This comprehensive approach addresses security concerns, affirming the protocol's readiness for practical deployment in cloud-assisted TMIS environments.

In [49], a lightweight mutual two-factor authentication mechanism is proposed for IoT devices and servers, leveraging PUFs and a hashing algorithm. The mechanism ensures secure authentication and session key agreement without storing cryptographic keys in non-volatile memory, thus addressing vulnerability concerns. The protocol is validated through formal analysis, demonstrating resilience against various attack scenarios while maintaining efficiency in terms of memory, server capacity, and energy consumption. By utilizing SRAM PUFs and an Arbiter PUF, the proposed mechanism achieves reliable two-factor mutual authentication without encryption, which can drain IoT device batteries quickly. Unlike existing methods requiring encryption, this approach relies solely on a hashing algorithm for mutual authentication, making it lightweight and suitable for power-constrained IoT networks.

In [50], a blockchain-assisted highly secure system for medical IoT devices is proposed using Lamport Merkle Digital Signature (LMDS). The Lamport Merkle Digital Signature Generation (LMDSG) model initially authenticates IoT devices by constructing a tree where leaves represent the hash function of sensitive patient medical data. A Centralized Healthcare Controller (CHC) determines the root of the LMDSG using Lamport Merkle Digital Signature Verification (LMDSV). In this verification, if the hash of the public key '*pbkey*' matches the leaf '*P gn*', the signature is valid. This LMDS technique efficiently detects malicious user behavior with minimal Computational Overhead (CO) and Computational Time (CT). Performance analysis considers CO, CT, and authentication accuracy, demonstrating higher security and lower CT and CO compared to existing methods in medical IoT systems.

In [51], a novel blockchain-based authentication scheme is designed to address IoT challenges. This framework leverages the modular square root algorithm integrated with blockchain technology to ensure an efficient authentication process. The security and effectiveness of the proposed scheme are demonstrated through comprehensive security analysis and detailed experiments. The proposed authentication scheme for IoT consists of four phases: system initialization, registration, authentication, and update/revocation. This scheme offers secure and lightweight authentication by combining blockchain with the MSR cryptographic algorithm, emphasizing decentralization, privacy preservation, and efficiency. A thorough security analysis evaluates the performance of the proposed scheme using Remix, comparing computation and communication costs with alternative methods to validate its effectiveness.

In [52], a blockchain-based decentralized authentication structure is proposed for IoT devices, organizing them into clusters based on computational capability, energy reserve, and location. Each cluster implements authentication through interconnected blockchains in a hierarchical manner. To reduce processing load, a consensus protocol verifies identity-based encryption key signatures of devices and their associated clusters. The proposed framework introduces a novel approach to authenticate IoT devices by grouping them into hierarchical clusters, each with its own blockchain for authentication. Clusters connect to a larger blockchain via a hash from the upper-level blockchain. A lightweight consensus algorithm validates nodes based on their public and cluster head key values, ensuring fast and efficient block validation within each cluster. This method maintains blockchain immutability while achieving speed and efficiency.

In [53], a lightweight authentication scheme using a consortium blockchain and a cryptocurrency-like digital token (LiIDCoin) is proposed to establish and manage trust among entities. LiIDCoin amounts manipulate the trust lifecycle. The scheme proves resilient against common attacks and is more efficient than competitors in terms of storage, communication, and authentication costs. It introduces cross-domain IoT authentication using a consortium blockchain and a novel data structure based on unspent transaction outputs (UTXO) with coin operations (issuance, transfer, query, revocation) for authentication and lifecycle management. LiIDCoin represents entity trust, demonstrated by transaction evidence on the blockchain, with the lifecycle managed by adjusting LiIDCoin amounts. Comprehensive security requirements are analyzed, and the scheme is implemented on the HLF platform, demonstrating superiority over competitors.

In [54], a three-factor authentication and key agreement protocol is introduced for Industrial Internet of Things (IIoT)

systems, leveraging Elliptic-Curve Cryptography (ECC). The protocol is designed to ensure both forward and backward secrecy by addressing the elliptic curve Diffie-Hellman problem. It can be applied to single-gateway scenarios and has also been extended to support multigateway environments. The proposed scheme incorporates three factors for authentication and utilizes ECC. It is specifically tailored for IIoT settings. The protocol's security is formally analyzed and proven to meet the required standards. By employing the computationally infeasible elliptic curve discrete logarithm problem, the proposed scheme achieves both forward and backward secrecy.

In [55], a three-factor authentication scheme called defense-in-depth is proposed for IoT environments on the blockchain. It employs mutual authentication and user authorization through smart card registration on a private blockchain, eliminating the need for a trusted server. The scheme integrates ECC for enhanced security. Security analysis, including assessments using the AVISPA tool, demonstrates the protocol's efficiency in terms of computational and communication costs. The protocol addresses vulnerabilities in IoT network authentication by implementing three-factor and mutual authentication, utilizing lightweight ECC cryptography to safeguard user privacy and enhance network resilience against security threats. Built on the blockchain platform, the protocol ensures data protection, decentralized management, transparency, and tamper-proof smart card creation for each user.

In [56], a three-factor authentication framework suitable for critical IoT applications is proposed. The framework incorporates identity, password, and a digital signature scheme. It utilizes a publish-subscribe pattern and leverages ECC and computationally low hash chains. The key features of the framework include mutual authentication of the Gateway node with both the remote user and sensor node, as well as the generation of dynamic session keys. Upon reviewing and analyzing relevant papers, it was found that none of the proposed protocols supported the user access level determination feature, which is a crucial security requirement in authentication protocols.

In [57], a method called PUFTAP-IoT proposed, which combines physical unclonable functions (PUF) and honey list techniques with three-factor authentication to design secure protocols for IoT environments. The aim is to resist attacks such as ID/password guessing, brute-force, and capture attacks. PUFTAP-IoT is analyzed for security using formal methods like BAN logic, Real-Or-Random (ROR) model, and scyther simulation tools. PUFTAP-IoT demonstrates its ability to provide secure services in IoT environments. The method incorporates PUF and honey list technologies to enhance security for sensing devices in the IoT environment, protecting against online guessing, brute-force attacks, and sensor takeover attacks.

In [58], a lightweight authenticated key agreement and access control protocol proposed for group communication in the blockchain using elliptic curve and bi-linear pairing. The protocol's secrecy is proven in the random-oracle paradigm, and a comprehensive heuristic security assessment is conducted to ensure its protection against potential threats and adherence to required security features. The protocol is utilized to implement the linear secret sharing (LSS) method. Non-transferable, unique assets such as user biometrics are employed for effective access control. The approach incorporates a robust login and authentication step, enabling quick identification of rogue users through appropriate threshold settings. The technique supports session key creation and authentication and combines access control and authentication into a single step.

In [59], the paper focuses on designing a secure user authentication scheme for cloud-assisted IoT systems. The proposed scheme is specifically designed for cloud-assisted IoT environments, with an emphasis on lightweight computation on gateways. It ensures secure access between remote users and IoT devices, incorporating desirable features such as forward secrecy and multi-factor security. The security of the scheme is rigorously proven using methods such as the random-oracle model, heuristic analysis, the ProVerif tool, and BAN logic. Additionally, the proposed scheme improves efficiency by offloading heavy computation and storage tasks to the cloud center. Overall, this paper presents a comprehensive and secure user authentication solution tailored for cloud-assisted IoT systems.

In [60], the Authenticated Devices Configuration Protocol (ADCP) is proposed to manage authentication and establish a secure overlay network within existing IoT networks. The Authenticated Device Transmission Protocol (ADTP) ensures secure communication within the overlay network. ADCP mitigates zero-day attacks and achieves zero round-trip-time key exchange. Both protocols use a distributed blockchain database optimized for data integrity to store authentication records, ensuring integrity. They are compatible with existing communication protocols and require no software reprogramming. Formal analysis confirms resilience against various attacks. This method addresses authentication-related security issues in IoT networks using blockchain, easily integrating into current networks. Experimental results show feasibility, and formal analysis confirms resilience against attacks, supported by a stochastic model showing security enhancement.

In [61], a hybrid blockchain-based many-to-many cross-domain authentication scheme is proposed for smart agriculture IoT networks. This scheme facilitates simultaneous mutual authentication between multiple devices and data service providers from various agricultural systems. It introduces a Groupable Batch Verification (GBV) algorithm that dynamically adjusts batch sizes to enhance cross-domain batch authentication flexibility. Additionally, the scheme includes a pseudonym update mechanism to safeguard device privacy and prevent illegal access by tracking malicious

devices. The proposed approach addresses certificate management and key escrow issues, offering cryptographic configuration adaptability. Security analysis and performance evaluation demonstrate practical security, efficiency, and affordability. The hybrid blockchain model reduces computational overhead and communication costs in many-to-many authentication scenarios, ensuring scalability and safety in cross-domain agricultural collaboration, unlike single-chain structures.

In [62], a novel lightweight authentication and key management scheme is proposed for IoT networks, integrating blockchain with Chebyshev chaotic maps. IoT devices undergo a registration process to obtain a temporary identity used for authentication and group key generation. During authentication, the device's temporary identity is updated and securely recorded on the blockchain, preventing exploitation by attackers. The Key Generation Center (KGC) uses Chebyshev polynomials to establish group keys without involving third parties, ensuring secure communication among group members. This approach guarantees efficient and secure group key generation and management, enhancing communication privacy within IoT networks. Formal and informal security analyses confirm the scheme's ability to meet rigorous security requirements while providing flexible key management. By integrating blockchain and Chebyshev chaotic maps, the proposed method delivers reliable and anonymous authentication alongside robust group key management for IoT devices.

In [63], an enhanced mutual authentication protocol is proposed for IoT-based Energy Internet (EI) using blockchain technology. The proposed protocol extends an existing smart grid authentication method by integrating blockchain-based security mechanisms to facilitate secure communication among IoT devices. To evaluate the protocol's performance, we conducted Caliper benchmarking and security testing using BAN logic and ProVerif. Experimental results demonstrate the protocol's achievement of both security and efficiency. Our blockchain-based solution enhances device authentication in IoT-based EI networks by utilizing a smart contract for user registration and verification. Multiple distributed registration authorities in the network improve resilience against attacks. This solution provides secure and efficient authentication for IoT devices in EI networks, validated through security analysis and performance evaluations with ProVerif, BAN logic, and the Caliper benchmark.

In [64], an authentication framework is proposed for an edge computing-enabled Internet of Things environment to establish secure communication between devices and edge servers, as well as among devices themselves. The protocol, named Device-Edge Authentication and Key Agreement (DEAKA), comprehensively addresses communication security. Additionally, a protocol called Device-Device Authentication and Key Agreement (DDAKA) is proposed for mutual authentication and key agreement among devices. The framework involves three entities: IoT devices, edge servers (ESs), and a trusted registration center (RC). Formal and informal security analyses demonstrate that the protocols meet a wide range of security requirements and can resist various security threats. Computational and transmission costs of the protocols are analyzed and compared, ensuring efficiency in resource utilization. This work extends existing authentication methods to cover inter-device communication in edge computing IoT environments, enhancing overall network security and reliability.

In [65], a blockchain-based secure remote authentication protocol (BSRA) proposed for fog-enabled Internet of Things systems. The protocol utilizes lightweight cryptographic primitives, including PUFs and cryptographic hash functions, to design an efficient authentication scheme. It incorporates temporary identities and authentication-piggybacking-synchronization techniques to ensure anonymity and effectiveness. The proposed protocol enables mutual authentication between users and IoT devices with the assistance of fog nodes, establishing distributed trust through blockchain technology. The scheme focuses on the use of computationally inexpensive cryptographic primitives for improved efficiency. Additionally, message synchronization is verified during the authentication process. Overall, the BSRA protocol offers a secure and efficient solution for remote authentication in fog-enabled IoT systems.

In Table 2, a summary of the advantages and disadvantages of the surveyed schemes is provided.

**Table 2.** Summary of surveyed schemes.

Ref	Advantage	Disadvantage
[45]	Lightweight scheme ensures secure message transmission, mutual authentication, and key agreement in blockchain-based WBANs.	Lack of formal security analysis, and potential scalability issues in large-scale WBAN deployments.
[46]	The proposed AKA protocols provide end-to-end security in IoT environments, addressing current security problems and meeting requirements.	The schemes have trade-offs, with Scheme 1 offering fast authentication but vulnerability to public key attacks, and Scheme 2 providing secure public key verification but slower performance.
[47]	Privacy-aware authentication protocol integrates PUFs and blockchain, providing security, resistance to attacks, and scalability for multi-server CE-IoT systems.	Protocol efficiency is moderate, with relatively long session key establishment and MC synchronization times for single requests.
[48]	The proposed protocol ensures data integrity, fine-grained access control, and security against various attacks in a cloud-assisted TMS environment.	The specific efficiency of the protocol is not provided, and the comparison with related protocols lacks detailed information.
[49]	Two-factor authentication protocol using hash functions, secure session key establishment, and robust defense against invasive attacks on IoT devices.	The specific efficiency and practical performance of the protocol are not provided, and further analysis is needed for different attack scenarios.
[50]	The proposed LMDS authentication technique for medical IoT systems reduces computational time, enhances security, and supports scalability.	The specific details of the security mechanisms and the potential limitations of the LMDS technique are not provided.
[51]	High security, Lightweight authentication system, Privacy preservation, Reduced computation costs, Decentralized system.	Complexity in implementation, Dependency on blockchain infrastructure, May require additional hardware resources, Reliance on MSR cryptographic algorithm.
[52]	Blockchain-based authentication framework reduces computational load, offers lightweight consensus, and enhances decentralization and efficiency for IoT devices.	The limitations of integrating the authentication values into smart contracts and the scalability of the proposed framework for a larger number of devices
[53]	Lightweight authentication scheme, Use of consortium blockchain for entity trust, LiIDCoin digital token for proving entity authenticity, Lifecycle management through manipulation of LiIDCoins, Satisfies security requirements	Limited to consortium blockchain, Dependency on the HLF platform, Limited analysis on real-world IoT applications, Limited application to cross-domain authentication scenarios, Privacy enhancement and fine-grained trust management not fully addressed
[54]	ECC-based authentication protocol, Suitable for single/multigateway scenarios, Achieves forward and backward secrecy, Efficient security attributes at reasonable computation cost	Limited to IIoT environment, ECC dependency for authentication, Limited real-world implementation analysis, Informal security analysis limitations
[55]	Efficient three-factor authentication protocol using a fuzzy extractor on the blockchain platform, providing security and privacy protection in heterogeneous IoT environments.	Additional complexity and overhead in terms of communication, computation, and storage requirements.
[56]	Signature-based 3-factor authentication using ECC and hash chains, Resistance to cryptographic attacks and formal security verification, Bandwidth and energy savings, reduced computing and communication costs	Dependency on publish-subscribe pattern and message queue telemetry transport, Potential limitations in scalability and adaptability to evolving IoT environments.
[57]	PUFTAP-IoT protocol addresses security vulnerabilities in IoT environments, provides secure mutual authentication.	High communication and storage overheads of PUFTAP-IoT in large-scale IoT deployments and different environments
[58]	Anonymous authenticated access control system for IoT group communication, Effective handling of access control with non-transferable, one-of-a-kind assets like user biometrics, Strong login and authentication step to quickly identify rogue users.	Limited discussion on scalability and adaptability to different IoT environments, Performance analysis and comparison research may not cover all aspects of system overhead improvement.
[59]	The proposed secure user authentication scheme for cloud-assisted IoT systems offers improved security, efficiency, and resource utilization.	The lack of scalability and resilience of the scheme to advanced attacks in diverse IoT environments
[60]	Authentication, security, data integrity, compatibility, zero-day attack mitigation, resilience.	Implementation complexity, potential resource overhead.
[61]	Addresses certificate management, key escrow, batch verification, pseudonym update, and device revocation, Practical security, efficiency, and affordability with low computational and communication costs.	Need to focused on the PBFT consensus algorithm without addressing other potential limitations, Limited discussion on the scalability and adaptability to non-agricultural IoT scenarios,
[62]	Quick authentication for new group managers, Secure against potential attacks, better security and functionality with lower computation cost and communication overhead.	Need to focused on optimizing group key generation and updating algorithm without addressing other potential limitations, Potential need for further optimization to reduce communication overhead in the proposed scheme.
[63]	Decentralized blockchain-based solution for device authentication in IoT-based EI networks, Resilient against attacks with distributed registration authorities, Easy integration with existing infrastructure and scalability.	The exploration of more sophisticated cryptographic primitives and integration with other blockchain-based solutions is needed. Further investigation is required for machine learning-based attack detection and scalability improvement through sharding.
[64]	Provably secure anonymous authentication for edge computing-enabled IoT, protecting against partial key-escrow	High communication cost, inability to counter ES impersonation attack, Limitation in scalability and adaptability to different IoT environments.



Ref	Advantage	Disadvantage
	attacks, leveraging blockchain, and demonstrating good security and performance.	
[65]	Blockchain-based authentication scheme for fog-enabled IoT, ensuring security even if a fog node is compromised, utilizing efficient cryptographic primitives.	Reliance on blockchain technology, additional computational and storage overhead lead to potentially affecting the overall performance and scalability of the system.

### 6. EVALUATION AND ANALYSIS

This section presents an analysis of various security requirements and vulnerabilities inherent in the surveyed schemes, along with an exploration of different methodologies employed. Furthermore, it provides an in-depth assessment of the efficacy and performance metrics of the evaluated methodologies.

In Table 3, we analyze the security aspects of the reviewed articles concerning anonymous authentication protocols within IoT platforms. Table 3 outlines the parameters utilized for evaluating the articles, including Anonymous and Unlinkable Sessions, Forward/Backward Security, Mutual Authentication, Untraceability, Data Verifiability, Key Agreement, and Scalability. Each article is scrutinized to determine its support for these security requirements. For instance, the extent to which it enables anonymous and unlinkable sessions, forward/backward security, mutual authentication, untraceability, data verifiability, key agreement, and scalability. Moreover, the evaluation examines the shortcomings of each article in meeting these security prerequisites. For example, some articles may excel in providing mutual authentication but fall short in ensuring untraceability. Through this comprehensive review, it becomes evident that no single article achieves complete coverage of all security and privacy requirements. Each article contributes differently to the overall security posture of IoT-based systems, with varying levels of support for the identified parameters. Consequently, the evaluation provides insights into the strengths and weaknesses of existing anonymous authentication protocols in IoT platforms, guiding future research directions for enhancing the security and privacy of such systems.

**Table 3.** Different security requirements in the surveyed schemes

Ref	F1	F2	F3	F4	F5	F6	F7
[45]	✓	✓	✓	-	-	-	-
[46]	-	-	✓	-	✓	-	-
[47]	✓	✓	✓	✓	-	-	✓
[48]	✓	✓	✓	✓	✓	-	✓
[49]	-	-	✓	-	-	-	✓
[50]	-	-	✓	-	✓	-	-
[51]	✓	-	✓	-	-	✓	✓
[52]	-	-	✓	-	-	-	✓
[53]	-	-	✓	-	-	-	✓
[54]	✓	✓	✓	✓	-	✓	-
[55]	✓	✓	✓	✓	✓	✓	-
[56]	✓	✓	✓	✓	-	-	-
[57]	✓	✓	✓	✓	-	-	-
[58]	✓	✓	✓	-	-	-	-
[59]	✓	✓	✓	✓	-	-	-
[60]	-	-	✓	-	✓	-	-
[61]	✓	✓	✓	-	-	✓	✓
[62]	✓	✓	✓	-	-	✓	-
[63]	-	✓	✓	-	-	-	✓
[64]	✓	✓	✓	-	-	✓	-
[65]	✓	✓	✓	✓	-	✓	-

F1=Anonymous and Unlinkable Sessions, F2=Forward/Backward Security, F3=Mutual Authentication, F4=Untraceability, F5= Data verifiability, F6=Key Agreement, F7=Scalability.

Table 4 provides a comprehensive overview of different attacks present in the surveyed schemes within the realm of surveyed authentication protocols for IoT platform. These attacks pose significant threats to the security and integrity of sensitive data and operations. The parameters evaluated in this table include Replay Attack, Impersonation/Capture Attack, Jamming/Desynchronization Attacks, Key Leakage, Machine Learning Attacks, Man-in-the-Middle Attack, Physical Attack, Denial of Service (DoS), Insider Attack, Password Exposure, and Decentralization. Replay Attack refers to the malicious act of intercepting and retransmitting data to gain unauthorized access or achieve other nefarious goals. Impersonation/Capture Attack involves an attacker posing as a legitimate entity to gain access to sensitive



information or perform unauthorized actions. Jamming/Desynchronization Attacks disrupt communication channels or synchronization processes, leading to system dysfunction or data manipulation. Key Leakage occurs when cryptographic keys are compromised, enabling unauthorized access to encrypted data. Machine Learning Attacks exploit vulnerabilities in machine learning algorithms or models to manipulate data or compromise system integrity. Man-in-the-Middle Attack intercepts communication between two parties to eavesdrop on or alter the exchanged data. Physical Attack involves the direct manipulation or tampering of hardware components to gain unauthorized access or disrupt system operations. DoS Attack floods the system with excessive traffic or requests, rendering it unable to fulfill legitimate requests. Insider Attack involves malicious actions by individuals with authorized access to the system, exploiting their privileges to compromise security. Password Exposure occurs when passwords or authentication credentials are exposed to unauthorized parties, leading to potential breaches. Decentralization refers to the distribution of system components or functions across multiple nodes, enhancing resilience against single points of failure or attacks. Each surveyed article addresses a combination of these attacks through various mechanisms and strategies tailored to the specific requirements and challenges of IoT-based systems. By comprehensively evaluating how each scheme tackles these threats, stakeholders can make informed decisions regarding the implementation of surveyed authentication protocols to safeguard data and operations. Table 4 serves as a valuable reference point for assessing the effectiveness and robustness of different approaches in mitigating security risks in IoT environments.

**Table 4.** Different attacks in the surveyed schemes

Ref	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
[45]	✓	✓	✓								
[46]	✓	✓		✓						✓	
[47]	✓	✓	✓	✓	✓						
[48]	✓	✓		✓		✓	✓		✓		✓
[49]	✓	✓		✓	✓	✓	✓	✓			
[50]											
[51]	✓	✓		✓		✓					
[52]		✓				✓		✓			✓
[53]	✓			✓		✓		✓			
[54]	✓	✓	✓				✓		✓		
[55]	✓	✓			✓	✓	✓	✓	✓		
[56]	✓	✓				✓	✓		✓		
[57]	✓	✓		✓		✓	✓		✓	✓	
[58]	✓	✓		✓		✓					✓
[59]		✓	✓				✓		✓	✓	
[60]		✓						✓		✓	✓
[61]	✓	✓				✓		✓			✓
[62]	✓	✓			✓	✓	✓		✓		
[63]	✓	✓			✓	✓		✓			✓
[64]	✓	✓				✓				✓	
[65]	✓	✓	✓		✓	✓	✓		✓	✓	

F1=Replay Attack, F2=Impersonation / Capture Attack, F3=Jamming/Desynchronization Attacks, F4=Key Leakage, F5=Machin Learning attacks, F6=Man-in-the middle attack, F7=Physical attack, F8=Denial of Service (DoS), F9=Insider attack, F10=Password Exposure, F11=Decentralization.

In Table 5, the evaluation of surveyed authentication protocols in IoT platforms is presented with a focus on Computational Costs and Communication Costs. These parameters are crucial considerations in assessing the practical feasibility and efficiency of implementing such protocols in real-world scenarios. Computational Costs refer to the amount of computational resources, such as processing power, required to execute the authentication protocols. Higher computational costs can impose significant overhead on IoT devices, which often have limited resources in terms of processing capabilities and energy consumption. Therefore, protocols with lower computational costs are generally preferred as they enable more efficient utilization of IoT resources and prolong device lifespan. Communication Costs, on the other hand, pertain to the amount of data exchanged between IoT devices and other components of the system during the authentication process. This includes both the volume of data transmitted and the frequency of communication, as excessive communication can lead to congestion, latency, and increased energy consumption in IoT networks. Protocols with lower communication costs optimize network bandwidth usage and reduce the burden on communication channels, enhancing overall system performance and scalability. In Table 5, each surveyed article is evaluated based on how it addresses and mitigates Computational Costs and Communication Costs within the context of authentication in IoT platforms. By comparing these parameters across different protocols, stakeholders can gauge the trade-offs between security, computational efficiency, and communication overhead. This comparative analysis

helps in identifying the most suitable protocol for specific deployment scenarios, considering factors such as device capabilities, network constraints, and security requirements. Furthermore, Table 5 serves as a valuable resource for researchers, developers, and decision-makers involved in the design and implementation of IoT-based systems, providing insights into the practical implications of various authentication protocols in terms of computational and communication costs. By understanding these costs, stakeholders can make informed decisions regarding protocol selection, deployment strategies, and resource allocation, ultimately ensuring the security and efficiency of IoT platforms in diverse application domains.

**Table 5.** Computational and Communication Costs

Ref	Computational cost	Communication cost
[45]	$10T_h + 8T_{XOR} + T_{sym}$	2880 bit
[46]	$6T_{EA} + 8T_{EM} + 4T_h$	
[47]	$8T_{GM} + 25T_h + 2T_{puf} + T_{MreDe} + T_{MreEn} + T_{Ga}$	3808 bit
[48]	$2T_{bp} + 13T_{mul} + 2T_{rng} + 9T_h$	3456 bit
[49]	$2T_{put} + 6T_{HMAC}$	1608 bit
[51]	$2T_{md} + 2T_{me} + T_{ae} + T_{ad}$	1888 bit
[53]		1467 bit
[54]	$35T_h + T_{fe} + 20T_{ecm} + 4T_{eca}$	4416 bit
[55]	$18T_h + 14T_x + 2T_{fe} + 2T_{ecm}$	1024 bit
[56]	$10T_{ecm} + 7T_h + 4T_{eca}$	2560 bit
[57]	$34T_h + 3T_{rg} + T_{puf} + 2T_{fe}$	1837 bit
[58]	$8T_h + 4T_{exp} + 2T_{bp}$	
[59]	$6T_{EM} + 31T_h + T_{fe}$	2720 bit
[60]	$6T_{EA} + 2T_{Em} + 5T_h$	2824 bit
[61]	$(9n + 3)T_{ecm} + (7n - 2)T_{eca} + 9nT_h$	4256 bit
[62]	$8T_h + 4T_c$	1056 bit
[63]	$6T_{EM}$	1408 bit
[64]	$4T_{EM} + T_{EA} + 5T_h + T_e$	3616 bit
[65]	$27T_h + T_{puf} + 2T_{sym}$	3680 bit

$T_{XOR}$ = XOR operation,  $T_{sym}$  = Symmetric encryption,  $T_{EA}$ : elliptic curve addition operation,  $T_{EM}$ : elliptic curve multiple operation,  $T_h$ : one-way hash function,  $T_{GM}$ =Scalar multiplication on G,  $T_{PUF}$ =PUF generation,  $T_{MreDe}$ =MRE decryption,  $T_{MreEn}$ =MRE encryption,  $T_{Ga}$ =Addition on G,  $T_{bp}$ =bilinear pairing operation,  $T_{mul}$ =scalar multiplication operation,  $T_{rng}$ =random number generation,  $T_{HMAC}$ =computing a hashed message authentication code,  $T_{me}$ =MSR encryption,  $T_{md}$ =MSR decryption,  $T_{ae}$ =AES encryption,  $T_{ad}$ =AES decryption,  $T_{fe}$ =Fuzzy extractor function,  $T_{ecm}$ = ECC point multiplication,  $T_{eca}$ = ECC point addition,  $T_{rg}$ =random nonce generation,  $T_{exp}$ =Modular Exponential Operation,  $T_c$ =Chebyshev mapping,  $T_e$ =Modular exponentiation.

## 7. CONCLUSION

In this paper, recently developed authentication techniques for IoT were surveyed. The analysis included a comprehensive comparison of these methods to highlight their respective strengths, weaknesses, and vulnerabilities against specific attacks. By understanding the distinct characteristics of each technique, we can better align them with the security requirements of IoT systems. Evaluation of computational and communication costs underscores the need for balancing security requirements with practical considerations such as resource constraints and network efficiency. Protocols that strike a judicious balance between security and performance emerge as promising candidates for real-

world deployment, offering scalable and efficient solutions for securing IoT platforms. Moreover, the analysis of attacks and countermeasures underscores the dynamic nature of security threats facing IoT systems, necessitating continuous innovation and adaptation in security protocols and practices. Looking forward, as IoT systems continue to proliferate and face escalating threats, there will be a pressing need to enhance existing authentication techniques. This will involve refining protocols, integrating new technologies like blockchain or biometrics, and implementing stronger encryption methods. Furthermore, ongoing modifications and advancements in authentication strategies will be essential to keep pace with evolving cyber threats and ensure the resilience and trustworthiness of IoT deployments in the future. Finally, future works should emphasize the development of standardized frameworks and testing environments to evaluate the effectiveness and interoperability of authentication protocols across diverse IoT ecosystems. This will facilitate the adoption of secure and scalable solutions capable of meeting the evolving demands of real-world deployments. Furthermore, interdisciplinary approaches that combine artificial intelligence and behavioral biometrics hold significant promise for enhancing real-time threat detection and user authentication. AI-driven models can analyze patterns and anomalies in user behavior, contributing to more dynamic and context-aware security measures. Additionally, exploring multi-factor authentication systems that blend traditional methods with novel biometric and environmental sensors can add layers of robustness to IoT security.

## REFERENCES

- [1] Abdul-Qawy, A. S., Pramod, P. J., Magesh, E., & Srinivasulu, T. (2015), "The internet of things (iot): An overview", *International Journal of Engineering Research and Applications*, 5(12), 71-82.
- [2] Kiamansouri, E., Barati, H., & Barati, A. (2022), "A two-level clustering based on fuzzy logic and content-based routing method in the internet of things", *Peer-to-Peer Networking and Applications*, 15(4), 2142-2159.
- [3] López, T. S., Ranasinghe, D. C., Patkai, B., & McFarlane, D. (2011), "Taxonomy, technology and applications of smart objects", *Information Systems Frontiers*, 13, 281-300.
- [4] Sharma, N., Shamkuwar, M., & Singh, I. (2019), "The history, present and future with IoT", *Internet of things and big data analytics for smart generation*, 27-51.
- [5] Akbari, M. R., Barati, H., & Barati, A. (2022), "An overlapping routing approach for sending data from things to the cloud inspired by fog technology in the large-scale IoT ecosystem", *Wireless Networks*, 28(2), 521-538.
- [6] Akbari, M. R., Barati, H., & Barati, A. (2022), "An efficient gray system theory-based routing protocol for energy consumption management in the Internet of Things using fog and cloud computing", *Computing*, 104(6), 1307-1335.
- [7] Shojarazavi, T., Barati, H., & Barati, A. (2022), "A wrapper method based on a modified two-step league championship algorithm for detecting botnets in IoT environments", *Computing*, 104(8), 1753-1774.
- [8] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013), "Internet of Things (IoT): A vision, architectural elements, and future directions", *Future generation computer systems*, 29(7), 1645-1660.
- [9] Onumanyi, A. J., Abu-Mahfouz, A. M., & Hancke, G. P. (2020), "Low power wide area network, cognitive radio and the Internet of Things: Potentials for integration", *Sensors*, 20(23), 6837.
- [10] Lee, I., & Lee, K. (2015), "The Internet of Things (IoT): Applications, investments, and challenges for enterprises", *Business horizons*, 58(4), 431-440.
- [11] Chataut, R., Phoummalayvane, A., & Akl, R. (2023), "Unleashing the power of IoT: A comprehensive review of IoT applications and future prospects in healthcare, agriculture, smart homes, smart cities, and industry 4.0", *Sensors*, 23(16), 7194.
- [12] Rehman, A., Saba, T., Kashif, M., Fati, S. M., Bahaj, S. A., & Chaudhry, H. (2022), "A revisit of internet of things technologies for monitoring and control strategies in smart agriculture", *Agronomy*, 12(1), 127.
- [13] Javed, A. R., Shahzad, F., ur Rehman, S., Zikria, Y. B., Razzak, I., Jalil, Z., & Xu, G. (2022), "Future smart cities: Requirements, emerging technologies, applications, challenges, and future aspects", *Cities*, 129, 103794.
- [14] Sun, P. J. (2019), "Privacy protection and data security in cloud computing: a survey, challenges, and solutions", *Ieee Access*, 7, 147420-147452.
- [15] Munirathinam, S. (2020), "Industry 4.0: Industrial internet of things (IIOT)" In *Advances in computers* (Vol. 117, No. 1, pp. 129-164). Elsevier.
- [16] Karale, A. (2021), "The challenges of IoT addressing security, ethics, privacy, and laws", *Internet of Things*, 15, 100420.
- [17] Obaid, O. I., & Salman, S. A. B. (2022), "Security and Privacy in IoT-based Healthcare Systems: A Review", *Mesopotamian Journal of Computer Science*, 2022, 29-39.
- [18] Chen, J. Q., & Benusa, A. (2017), "HIPAA security compliance challenges: The case for small healthcare providers", *International Journal of Healthcare Management*, 10(2), 135-146.
- [19] Sun, Y., Lo, F. P. W., & Lo, B. (2019), "Security and privacy for the internet of medical things enabled healthcare systems: A survey", *IEEE Access*, 7, 183339-183355.
- [20] Hasan, M. K., Ghazal, T. M., Saeed, R. A., Pandey, B., Gohel, H., Eshmawi, A. A., ... & Alkhasawneh, H. M. (2022), "A review on security threats, vulnerabilities, and counter measures of 5G enabled Internet-of-Medical-Things", *IET communications*, 16(5), 421-432.
- [21] Hamidi, H. (2019), "An approach to develop the smart health using Internet of Things and authentication based on biometric technology", *Future generation computer systems*, 91, 434-449.

- [22] Newaz, A. I., Sikder, A. K., Rahman, M. A., & Uluagac, A. S. (2021), "A survey on security and privacy issues in modern healthcare systems: Attacks and defenses", *ACM Transactions on Computing for Healthcare*, 2(3), 1-44.
- [23] Nauman, A., Qadri, Y. A., Amjad, M., Zikria, Y. B., Afzal, M. K., & Kim, S. W. (2020), "Multimedia Internet of Things: A comprehensive survey", *Ieee Access*, 8, 8202-8250.
- [24] El-Hajj, M., Fadlallah, A., Chamoun, M., & Serhrouchni, A. (2019), "A survey of internet of things (IoT) authentication schemes", *Sensors*, 19(5), 1141.
- [25] Rajakumari, S., Azhagumeena, S., Devi, A. B., & Ananthi, M. (2017, February), "Upgraded living think-IoT and big data", In *2017 2nd International Conference on Computing and Communications Technologies (ICCT)* (pp. 181-184). IEEE.
- [26] Gupta, B. B., & Quamara, M. (2020), "An overview of Internet of Things (IoT): Architectural aspects, challenges, and protocols", *Concurrency and Computation: Practice and Experience*, 32(21), e4946.
- [27] Ray, P. P. (2018), "A survey on Internet of Things architectures", *Journal of King Saud University-Computer and Information Sciences*, 30(3), 291-319.
- [28] Yaqoob, T., Abbas, H., & Atiquzzaman, M. (2019), "Security vulnerabilities, attacks, countermeasures, and regulations of networked medical devices—A review", *IEEE Communications Surveys & Tutorials*, 21(4), 3723-3768.
- [29] Riggs, H., Tufail, S., Parvez, I., Tariq, M., Khan, M. A., Amir, A., ... & Sarwat, A. I. (2023), "Impact, vulnerabilities, and mitigation strategies for cyber-secure critical infrastructure", *Sensors*, 23(8), 4060.
- [30] Mohanty, J., Mishra, S., Patra, S., Pati, B., & Panigrahi, C. R. (2021), "IoT security, challenges, and solutions: a review", *Progress in Advanced Computing and Intelligent Engineering: Proceedings of ICACIE 2019, Volume 2*, 493-504.
- [31] Khattak, H. A., Shah, M. A., Khan, S., Ali, I., & Imran, M. (2019), "Perception layer security in Internet of Things", *Future Generation Computer Systems*, 100, 144-164.
- [32] Jing, Q., Vasilakos, A. V., Wan, J., Lu, J., & Qiu, D. (2014), "Security of the Internet of Things: perspectives and challenges", *Wireless networks*, 20, 2481-2501.
- [33] Nastase, L. (2017, May), "Security in the internet of things: A survey on application layer protocols", In *2017 21st international conference on control systems and computer science (CSCS)* (pp. 659-666). IEEE.
- [34] Tewari, A., & Gupta, B. B. (2020), "Security, privacy and trust of different layers in Internet-of-Things (IoTs) framework", *Future generation computer systems*, 108, 909-920.
- [35] Ngu, A. H., Gutierrez, M., Metsis, V., Nepal, S., & Sheng, Q. Z. (2016), "IoT middleware: A survey on issues and enabling technologies", *IEEE Internet of Things Journal*, 4(1), 1-20.
- [36] Chaqfeh, M. A., & Mohamed, N. (2012, May), "Challenges in middleware solutions for the internet of things", In *2012 international conference on collaboration technologies and systems (CTS)* (pp. 21-26). IEEE.
- [37] Nebbione, G., & Calzarossa, M. C. (2020), "Security of IoT application layer protocols: Challenges and findings", *Future Internet*, 12(3), 55.
- [38] Karagiannis, V., Chatzimisios, P., Vazquez-Gallego, F., & Alonso-Zarate, J. (2015), "A survey on application layer protocols for the internet of things", *Transaction on IoT and Cloud computing*, 3(1), 11-17.
- [39] Deep, S., Zheng, X., Jolfaei, A., Yu, D., Ostovari, P., & Kashif Bashir, A. (2022), "A survey of security and privacy issues in the Internet of Things from the layered context", *Transactions on Emerging Telecommunications Technologies*, 33(6), e3935.
- [40] Cresitello-Dittmar, B. (2016), "Application of the blockchain for authentication and verification of identity", *Independent Paper*.
- [41] Chenchev, I., Aleksieva-Petrova, A., & Petrov, M. (2021), "Authentication Mechanisms and Classification: A Literature Survey", In *Intelligent Computing: Proceedings of the 2021 Computing Conference, Volume 3* (pp. 1051-1070). Springer International Publishing.
- [42] Patwary, A. A. N., Naha, R. K., Garg, S., Battula, S. K., Patwary, M. A. K., Aghasian, E., ... & Gong, M. (2021), "Towards secure fog computing: A survey on trust management, privacy, authentication, threats and access control", *Electronics*, 10(10), 1171.
- [43] Nandy, T., Idris, M. Y. I. B., Noor, R. M., Kiah, L. M., Lun, L. S., Juma'at, N. B. A., ... & Bhattacharyya, S. (2019), "Review on security of internet of things authentication mechanism", *IEEE Access*, 7, 151054-151089.
- [44] Ferrag, M. A., Maglaras, L. A., Janicke, H., Jiang, J., & Shu, L. (2017), "Authentication protocols for internet of things: a comprehensive survey", *Security and Communication Networks*, 2017.
- [45] Xu, J., Meng, X., Liang, W., Peng, L., Xu, Z., & Li, K. C. (2020), "A hybrid mutual authentication scheme based on blockchain technology for WBANs", In *Blockchain and Trustworthy Systems: First International Conference, BlockSys 2019, Guangzhou, China, December 7–8, 2019, Proceedings 1* (pp. 350-362). Springer Singapore.
- [46] Lee, D. H., & Lee, I. Y. (2020), "A lightweight authentication and key agreement schemes for IoT environments", *Sensors*, 20(18), 5350.
- [47] Zhang, Y., Li, B., Liu, B., Hu, Y., & Zheng, H. (2021), "A privacy-aware PUFs-based multiserver authentication protocol in cloud-edge IoT systems using blockchain", *IEEE Internet of Things Journal*, 8(18), 13958-13974.

- [48] Son, S., Lee, J., Kim, M., Yu, S., Das, A. K., & Park, Y. (2020), “**Design of secure authentication protocol for cloud-assisted telecare medical information system using blockchain**”, *IEEE Access*, 8, 192177-192191.
- [49] Mostafa, A., Lee, S. J., & Peker, Y. K. (2020), “Physical unclonable function and hashing are all you need to mutually authenticate iot devices”, *Sensors*, 20(16), 4361.
- [50] Alzubi, J. A. (2021), “**Blockchain-based Lamport Merkle digital signature: authentication tool in IoT healthcare**”, *Computer Communications*, 170, 200-208.
- [51] Yang, X., Yang, X., Yi, X., Khalil, I., Zhou, X., He, D., ... & Nepal, S. (2021), “**Blockchain-based secure and lightweight authentication for Internet of Things**”, *IEEE Internet of Things Journal*, 9(5), 3321-3332.
- [52] Al Ahmed, M. T., Hashim, F., Hashim, S. J., & Abdullah, A. (2022), “**Hierarchical blockchain structure for node authentication in IoT networks**”, *Egyptian Informatics Journal*, 23(2), 345-361.
- [53] Zhang, Y., Luo, Y., Chen, X., Tong, F., Xu, Y., Tao, J., & Cheng, G. (2022), “**A lightweight authentication scheme based on consortium blockchain for cross-domain IoT**”, *Security and Communication Networks*, 2022, 1-15.
- [54] Zhao, X., Li, D., & Li, H. (2022), “**Practical three-factor authentication protocol based on elliptic curve cryptography for industrial internet of things**”, *Sensors*, 22(19), 7510.
- [55] Mirsarai, A. G., Barati, A., & Barati, H. (2022), “**A secure three-factor authentication scheme for IoT environments**”, *Journal of Parallel and Distributed Computing*, 169, 87-105.
- [56] Saqib, M., Jasra, B., & Moon, A. H. (2022), “**A lightweight three factor authentication framework for IoT based critical applications**”, *Journal of King Saud University-Computer and Information Sciences*, 34(9), 6925-6937.
- [57] Lee, J., Oh, J., Kwon, D., Kim, M., Yu, S., Jho, N. S., & Park, Y. (2022), “**PUFTAP-IoT: PUF-based three-factor authentication protocol in IoT environment focused on sensing devices**”, *Sensors*, 22(18), 7075.
- [58] Singh, A., Chandra, H., Rana, S., & Chhikara, D. (2023), “**Blockchain based authentication and access control protocol for IoT**”, *Multimedia Tools and Applications*, 1-23.
- [59] Wang, C., Wang, D., Duan, Y., & Tao, X. (2023), “**Secure and lightweight user authentication scheme for cloud-assisted internet of things**”, *IEEE Transactions on Information Forensics and Security*.
- [60] Lau, C. H., Yeung, K. H., Yan, F., & Chan, S. (2023), “**Blockchain-based authentication and secure communication in IoT networks**”, *Security and Privacy*, 6(6), e319.
- [61] Luo, F., Huang, R., & Xie, Y. (2024), “**Hybrid blockchain-based many-to-many cross-domain authentication scheme for smart agriculture IoT networks**”, *Journal of King Saud University-Computer and Information Sciences*, 101946.
- [62] Long, Y., Peng, C., Tan, W., & Chen, Y. (2024), “**Blockchain-Based Anonymous Authentication and Key Management for Internet of Things With Chebyshev Chaotic Maps**”, *IEEE Transactions on Industrial Informatics*.
- [63] Benrebbouh, C., Mansouri, H., Cherbal, S., & Pathan, A. S. K. (2024), “**Enhanced secure and efficient mutual authentication protocol in IoT-based energy internet using blockchain**”, *Peer-to-Peer Networking and Applications*, 17(1), 68-88.
- [64] Zhang, S., & Cao, D. (2024), “**A blockchain-based provably secure anonymous authentication for edge computing-enabled IoT**”, *The Journal of Supercomputing*, 80(5), 6778-6808.
- [65] Guo, Y., Zhang, Z., Guo, Y., & Xiong, P. (2023), “**BSRA: Blockchain-based secure remote authentication scheme for the fog-enabled Internet of Things**”, *IEEE Internet of Things Journal*.

