Systematic Literature Review of AI/ML in Software-Defined Networks Using the Snowballing Approach

João C. Ferreira[†], Daniel Teixeira[0000−0002−9596−5598]*[†], and Joaquim Macedo[0000−0002−5892−1289]

University of Minho, 4710-057 Braga, Portugal
{pg42858,pg42857}@alunos.uminho.pt, macedo@di.uminho.pt

Abstract. Current networks need to host an array of heterogeneous devices with different resource requirements and traffic outputs while maintaining acceptable QoS. To meet these requirements, networks have become increasingly more complex and difficult to manage, configure and monitor. To make networks more easily manageable and controllable, researchers and operators proposed to use software programs that can monitor the network and configure it on-demand automatically. With Software-Defined Networks, we can build programs to efficiently manage the network through intelligent algorithms. In this study we conducted a systematic analysis focused on the use of AI/ML algorithms to improve SDN functions. We used a snowballing approach to organize and select articles to review. Following the analysis of 1200 articles (and the acceptance of 38), we present an overview of the state-of-the-art.

Keywords: Software-Defined Networks · Artificial Intelligence · Machine Learning · Systematic Literature Review

1 Introduction

The increased service heterogeneity and consumption reveal the need for new and efficient network architectures. Current networks are configured with complex static rules to satisfy Service-Level Agreements (SLAs) and maintain quality of service (QoS). In multi-vendor environments, the management and configuration operations tend to increase in complexity. Traditional network-centric architectures are unable to solve these challenges. Hence, the network-centric gives way to application-centric paradigms with Software-Defined Networks (SDN) and network programmability and automation [14].

The SDN architecture consists of removing the control plane functions from the standard networking devices and transferring them to a centralized controller. The SDN controller hosts APIs to provide access and to program the network resources. However, through these APIs the programs must efficiently

*Corresponding author: Daniel Teixeira, pg42857@alunos.uminho.pt
†These authors contributed equally to this paper.

optimize network performance and security, and artificial intelligence (AI) and machine learning (ML) algorithms can help achieve this goal.

Researchers and industry experts have carried out multiple systematic reviews of SDN studies. The authors in [7] report on various techniques used in SDN to ensure load balancing, some of which include AI algorithms. In [30], Ray et al. present a study of IoT devices in SDN architectures. Also discussed in [36] is the challenge of implementing AI/ML in SDNs, presenting articles that explore this theme and creating an overview of what they try to propose and explain.

Although these reviews present very interesting perspectives and a deep analysis on the topic, none address the application of AI/ML to SDNs as a whole, focusing only on specific aspects such as intrusion detection systems and load balancing. Moreover, no articles were found that followed a systematic literature review approach with a strict and transparent selection system such as the snowballing approach. Hence, the current study mainly attempts to (i) apply the snowballing approach and collect a set of articles with a rigorous process; (ii) discuss how AI/ML can improve performance and solve certain issues in SDNs.

In Section II we describe the methodology of the study, encompassing the main research questions and considered criteria. We present the results in Section III and discuss them in view of the questions. Section IV closes the paper with some final remarks.

2 Methodology

The goal of this study is to gather a set of articles and review them to answer multiple research questions. The posed research questions are:
RQ1. What kind of AI/ML mechanisms are applied to SDN?
RQ2. Does AI/ML positively influence the performance of SDNs?
RQ3. What are the main limitations of using AI/ML in SDNs?

We used the snowballing method [40] to collect a set of papers to review. The process starts with a set of initial papers that are analyzed following inclusion/exclusion criteria. From those included, their references and citing articles are gathered and analyzed in a following iteration. The snowballing process is complete when no article is accepted in an iteration.

To select an appropriate start set, the terms used in our search were Software-Defined Networks, Artificial Intelligence, Machine Learning, and the respective acronyms: *("SDN" OR "Software Defined Network") AND ("Artificial Intelligence" OR "AI" OR "Machine Learning" OR "ML")*

2.1 Start Set and Criteria

To construct the start set we introduced the search string in the Google Scholar, Core, and IEEE Xplore search engines. We then chose the first five papers from each.

Before analyzing the papers, we defined the following acceptance criteria:

- Publish date between 2011 and 2020 (release date of OpenFlow);
- Written in English;
- Peer-reviewed and published in first or second-quartile journals (Scimago ranking);
- Report directly to the topics at hand, i.e., AI/ML applied to SDN;
- Articles to which the authors have been granted access.

With the acceptance criteria, we selected a set of articles from reliable sources and that directly explore AI/ML in SDNs. Ultimately, from the fifteen articles of the candidate starting set, only six respected the criteria: [41, 11, 29, 22, 27, 36].

2.2 Iterations

Table 1 illustrates the snowballing process by presenting the number of analysed and accepted papers in each phase. In total, 259 were excluded based on year of publishing, 88 were not published in sufficiently high ranked journals or conferences, 9 were inaccessible, 284 were duplicates, 1 was not an article, and 576 did not report directly to the topics at hand.

From this process, carried out between November and December 2020, we selected a total of 38 articles.

Table 1. Iteration results with number of analysed and accepted papers for each snowballing step.

	Start Set	First		Second		Third	
		Backward	**Forward**	**Backward**	**Forward**	**Backward**	**Forward**
Analysed	15	412	212	253	161	129	60
Accepted	6	20	4	2	6	0	0

3 Discussion

This section focuses on presenting and discussing our findings, and giving an overview over current and future trends.

3.1 Applied AI/ML algorithms

Table 2 shows the use of AI/ML algorithms by the articles. The classification demonstrates that neural networks (NNs) algorithms are the most popular, followed by other supervised learning algorithms. Other mechanisms have been presented in several studies, however, they were not as widely used, such as the self-organising maps.

The dominance of supervised learning methods over the rest can be explained by the popularity of deep learning and neural networks, as well as DT and RF algorithms. Supervised learning requires labelled data to optimize model parameters, while unsupervised learning learns patterns from unlabelled data.

Table 2. AI/ML algorithms found in the accepted papers.

AI/ML Algorithm		**Articles**
K-Nearest Neighbours (KNN)		[41, 11, 22, 27, 1, 39, 37, 26, 38, 28, 2]
Decision Tree (DT)		[41, 29, 22, 27, 38, 28, 1, 2, 8, 17, 16, 4, 21]
Random Forest (RF)		[36, 41, 22, 28, 2, 4, 3, 20, 9]
Neural Network (NN)	General	[41, 22, 26, 38, 2, 8, 17, 4, 20, 23, 32, 25, 6] [31, 15, 34, 35, 13, 42, 33]
	Deep NN	[41, 22, 1, 38, 3, 23, 31, 13, 10]
	Convolutional NN	[36, 41, 22, 9, 5, 24]
	Recurrent NN	[41, 22, 20, 9, 6]
Reinforcement Learning (RL)	General	[41, 22, 20, 18]
	Deep RL	[41, 22, 1, 15, 13, 42, 33, 24, 19]
	RL-Based Game Theory	[41]
Self-Organising Map		[36, 41, 22]

In this sense, supervised learning also seems to be more easily applied to improve network decision making in, e.g., routing and QoS.

Figure 1, depicts the use of AI/ML algorithms per year, to demonstrate the evolution in research trends. Again supervised learning algorithms prevail over both unsupervised and RL, being characterized by a peak on 2017 and a variable decline in following years. Interestingly, RL has seen a slow and steady increase. This reveals a promising prospect for the use of SDNs in IoT, vehicular networking, and 5G access networks, which constitute dynamic environments where RL is known to outperform supervised and unsupervised learning approaches. The use of RL to teach the network and SDN controllers how to adapt to changes in traffic and resource demand is expected to increase based on these results, specially with the introduction of SDNs in more complex networks.

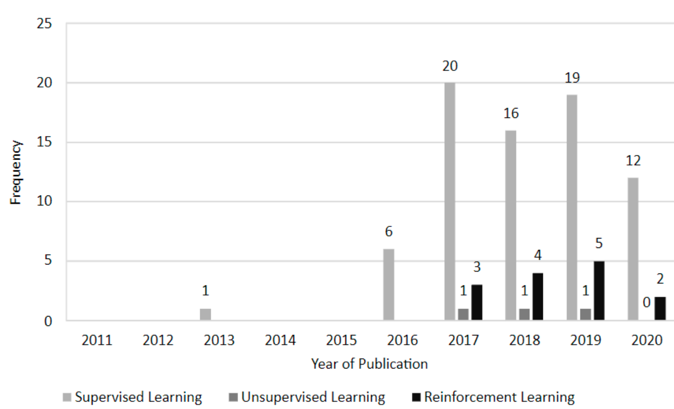

Fig. 1. Types of AI/ML algorithms by article's year of publication.

3.2 Impact of AI/ML on SDNs

Table 3 outlines the effect of AI/ML in SDNs: AI helps to improve QoS, enhances autonomous network managing and controller, enables intelligent optimization of resources, and increases security. Although most mechanisms are proof of concepts, they clearly demonstrate the functionality gain of AI/ML, particularly in managing QoS and QoE, and network automation.

These findings suggest that AI/ML algorithms can be successfully applied to SDNs with potential to accommodate current and future service requirements.

Table 3. Impact of AI/ML algorithms in the accepted papers.

Use Cases	Articles
Optimise network resource usage	[2, 3, 8, 9, 12, 16, 19, 23, 41]
Enhance autonomous network management and configuration	[3, 5, 8, 11–13, 15, 18, 17, 20, 22, 23, 25, 24, 31–34, 37, 42]
Monitor load balancing	[11, 22, 38, 17]
Improve network security	[36, 29, 22, 27, 26, 6, 35, 10]
Guarantee Quality of Service	[39, 37, 28, 16, 4, 21, 9]

3.3 Limitations of applying AI/ML in SDNs

Although AI/ML methods help in many fields, there is a greater truth: there is no free lunch. In SDNs the same axiom applies, as illustrated in Table 4. A large percentage of studies report difficulties in the implementation of the AI/ML mechanisms (71%) and there's definitely room for improvement. The main difficulties are (i) learning distortion, (ii) difficulty in finding good training sets, (iii) and processing of high data flows without sampling.

Table 4. Limitations of AI/ML algorithms in SDN environments.

AI/ML Limitations	Articles
Obtaining high-quality labelled data	[41, 22, 27, 38, 23, 25, 42, 18, 12]
Distorted learning	[11, 22, 39, 26, 21, 20, 25, 34, 24]
Difficulty in collecting high-speed traffic	[36, 37, 2, 3, 9, 31, 15, 13, 5, 19]
Implementation cost and expertise	[36, 2]
Change of mindset by engineers and IT professionals	[25]

4 Conclusion

We reviewed 38 articles (from a pool of 1200) to study the use of AI/ML techniques in SDNs. The results suggest that supervised learning strongly prevails over unsupervised learning and reinforcement learning algorithms. Most studies find that NNs are the best solution to optimise SDNs and enhance intelligence. However, RL has seen a slight increase in adherence and may start to see a greater increase in solving SDN problems in environments where multiple heterogeneous devices compete for network resources (e.g., 5G networks). Both supervised and reinforcement learning improve network management, automation, network performance and QoS. Although the lack of quality datasets and learning distortion are concerning aspects, these are certain to improve with future research.

The exploitation of AI/ML in SDN applications improves these architecture's potential and value in the academia and industry. We predict that future trends will focus even more upon AI/ML techniques since they introduce considerable performance gains; studies will move from proof-of-concept designs to solid real-world setups as researchers and industry experts turn from design to implementation.

References

1. Abar, T., Ben Letaifa, A., El Asmi, S.: Machine learning based QoE prediction in SDN networks. IWCMC 2017 pp. 1395–1400 (jul 2017). https://doi.org/10.1109/IWCMC.2017.7986488
2. Abar, T., Ben Letaifa, A., El Asmi, S.: Enhancing QoE based on machine learning and dash in SDN networks. WAINA 2018 pp. 258–263 (jul 2018). https://doi.org/10.1109/WAINA.2018.00095
3. Alvizu, R., Troia, S., Maier, G., Pattavina, A.: Matheuristic with machine-learning-based prediction for software-defined mobile metro-core networks. Journal of Optical Communications and Networking **9**(9), D19–D30 (sep 2017). https://doi.org/10.1364/JOCN.9.000D19
4. Amaral, P., Dinis, J., Pinto, P., Bernardo, L., Tavares, J., Mamede, H.S.: Machine learning in software defined networks: Data collection and traffic classification. ICNP 2016 pp. 91–95 (dec 2016). https://doi.org/10.1109/ICNP.2016.7785327
5. Azzouni, A., Boutaba, R., Pujolle, G.: NeuRoute: Predictive dynamic routing for software-defined networks. CNSM 2017 pp. 1–6 (nov 2017). https://doi.org/10.23919/CNSM.2017.8256059
6. Azzouni, A., Pujolle, G.: NeuTM: A neural network-based framework for traffic matrix prediction in SDN. NOMS 2018 pp. 1–5 (jul 2018). https://doi.org/10.1109/NOMS.2018.8406199
7. Belgaum, M.R., Musa, S., Alam, M.M., Su'Ud, M.M.: A Systematic Review of Load Balancing Techniques in Software-Defined Networking. IEEE Access **8**, 98612–98636 (2020). https://doi.org/10.1109/ACCESS.2020.2995849
8. Ben Letaifa, A.: Adaptive QoE monitoring architecture in SDN networks: Video streaming services case. IWCMC 2017 pp. 1383–1388 (jul 2017). https://doi.org/10.1109/IWCMC.2017.7986486
9. Bouraqia, K., Sabir, E., Sadik, M., Ladid, L.: Quality of Experience for Streaming Services: Measurements, Challenges and Insights. IEEE Access **8**, 13341–13361 (2020). https://doi.org/10.1109/ACCESS.2020.2965099
10. Budhraja, K.K., Malvankar, A., Bahrami, M., Kundu, C., Kundu, A., Singhal, M.: Risk-Based Packet Routing for Privacy and Compliance-Preserving SDN. CLOUD 2017 pp. 761–765 (sep 2017). https://doi.org/10.1109/CLOUD.2017.109
11. Cao, Y., Wang, R., Chen, M., Barnawi, A.: AI Agent in Software-Defined Network: Agent-Based Network Service Prediction and Wireless Resource Scheduling Optimization. IEEE Internet of Things Journal **7**(7), 5816–5826 (jul 2020). https://doi.org/10.1109/JIOT.2019.2950730

12. Chemouil, P., Hui, P., Kellerer, W., Li, Y., Stadler, R., Wen, Y., Zhang, Y.: Special Issue on Artificial Intelligence and Machine Learning for Networking and Communications. IEEE Journal on Selected Areas in Communications **37**(6), 1185–1191 (jun 2019). https://doi.org/10.1109/JSAC.2019.2909076
13. Chen, J., Xiao, Z., Xing, H., Dai, P., Luo, S., Iqbal, M.A.: STDPG: A spatio-temporal deterministic policy gradient agent for dynamic routing in SDN. CoRR (apr 2020), https://arxiv.org/abs/2004.09783
14. Cisco: The Art of Application-Centric Networking. Tech. rep., https://www.cisco.com/c/dam/en/us/solutions/collateral/borderless-networks/officeextend-solution/cisco_td_030513_fin.pdf
15. Glick, M., Rastegarfar, H.: Scheduling and control in hybrid data centers. SUM 2017 pp. 115–116 (aug 2017). https://doi.org/10.1109/PHOSST.2017.8012677
16. Hamdan, M., Mohammed, B., Humayun, U., Abdelaziz, A., Khan, S., Ali, M.A., Imran, M., Marsono, M.N.: Flow-Aware Elephant Flow Detection for Software-Defined Networks. IEEE Access **8**, 72585–72597 (2020). https://doi.org/10.1109/ACCESS.2020.2987977
17. He, M., Kalmbach, P., Blenk, A., Kellerer, W., Schmid, S.: Algorithm-data driven optimization of adaptive communication networks. ICNP 2017 (nov 2017). https://doi.org/10.1109/ICNP.2017.8117592
18. He, Y., Richard Yu, F., Zhao, N., Leung, V.C., Yin, H.: Software-Defined Networks with Mobile Edge Computing and Caching for Smart Cities: A Big Data Deep Reinforcement Learning Approach. IEEE Communications Magazine **55**(12), 31–37 (dec 2017). https://doi.org/10.1109/MCOM.2017.1700246
19. Huang, X., Yuan, T., Qiao, G., Ren, Y.: Deep Reinforcement Learning for Multimedia Traffic Control in Software Defined Networking. IEEE Network **32**(6), 35–41 (nov 2018). https://doi.org/10.1109/MNET.2018.1800097
20. Kim, H.G., Lee, D.Y., Jeong, S.Y., Choi, H., Yoo, J.H., Hong, J.W.K.: Machine learning-based method for prediction of virtual network function resource demands. NetSoft 2019 pp. 405–413 (jun 2019). https://doi.org/10.1109/NETSOFT.2019.8806687
21. Kolomvatsos, K., Anagnostopoulos, C., Marnerides, A.K., Ni, Q., Hadjiefthymiades, S., Pezaros, D.P.: Uncertainty-driven ensemble forecasting of QoS in Software Defined Networks. ISCC 2017 pp. 1284–1289 (sep 2017). https://doi.org/10.1109/ISCC.2017.8024701
22. Latah, M., Toker, L.: Artificial Intelligence Enabled Software Defined Networking: A Comprehensive Overview. IET Networks **8**(2), 79–99 (mar 2018). https://doi.org/10.1049/iet-net.2018.5082
23. Liu, W.X., Zhang, J., Liang, Z.W., Peng, L.X., Cai, J.: Content Popularity Prediction and Caching for ICN: A Deep Learning Approach with SDN. IEEE Access **6**, 5075–5089 (dec 2017). https://doi.org/10.1109/ACCESS.2017.2781716
24. Mao, B., Tang, F., Fadlullah, Z.M., Kato, N.: An Intelligent Route Computation Approach Based on Real-Time Deep Learning Strategy for Software Defined Communication Systems. IEEE Transactions on Emerging Topics in Computing **9**(3), 1554–1565 (feb 2019). https://doi.org/10.1109/TETC.2019.2899407
25. Mestres, A., Rodriguez-Natal, A., Carner, J., Barlet-Ros, P., Alarcón, E., Solé, M., Muntés-Mulero, V., Meyer, D., Barkai, S., J., Hibbett, M., Estrada, G., Ma'ruf, K., Coras, F., Ermagan, V., Latapie, H., Cassar, C., Evans, J., Maino, F., Walrand, J., Cabellos, A.: Knowledge-Defined Networking. SIGCOMM Comput. Commun. Rev. **47**(3), 2–10 (sep 2017). https://doi.org/10.1145/3138808.3138810

26. Nam, T.M., Phong, P.H., Khoa, T.D., Huong, T.T., Nam, P.N., Thanh, N.H., Thang, L.X., Tuan, P.A., Dung, L.Q., Loi, V.D.: Self-organizing map-based approaches in DDoS flooding detection using SDN. Int'l Conf. on Information Networking pp. 249–254 (apr 2018). https://doi.org/10.1109/ICOIN.2018.8343119
27. Nanda, S., Zafari, F., Decusatis, C., Wedaa, E., Yang, B.: Predicting network attack patterns in SDN using machine learning approach. NFV-SDN 2016 pp. 167–172 (may 2017). https://doi.org/10.1109/NFV-SDN.2016.7919493
28. Owusu, A.I., Nayak, A.: An Intelligent Traffic Classification in SDN-IoT: A Machine Learning Approach. BlackSeaCom 2020 (may 2020). https://doi.org/10.1109/BLACKSEACOM48709.2020.9235019
29. Qazi, Z.A., Lee, J., Jin, T., Bellala, G., Arndt, M., Noubir, G.: Application-awareness in SDN. SIGCOMM 2013 pp. 487–488 (2013). https://doi.org/10.1145/2486001.2491700
30. Ray, P.P., Kumar, N.: SDN/NFV architectures for edge-cloud oriented IoT: A systematic review. Computer Communications **169**, 129–153 (mar 2021). https://doi.org/10.1016/J.COMCOM.2021.01.018
31. Rischke, J., Sossalla, P., Salah, H., Fitzek, F.H., Reisslein, M.: QR-SDN: Towards reinforcement learning states, actions, and rewards for direct flow routing in software-defined networks. IEEE Access **8**, 174773–174791 (2020). https://doi.org/10.1109/ACCESS.2020.3025432
32. Sendra, S., Rego, A., Lloret, J., Jimenez, J.M., Romero, O.: Including artificial intelligence in a routing protocol using Software Defined Networks. ICC Workshops 2017 pp. 670–674 (jun 2017). https://doi.org/10.1109/ICCW.2017.7962735
33. Shreya, T., Mulla, M.M., Shinde, S., Narayan, D.G.: Ant Colony Optimization-based Dynamic Routing in Software Defined Networks. ICCCNT 2020 (jul 2020). https://doi.org/10.1109/ICCCNT49239.2020.9225287
34. Singh, S.: SDN (Software Defined Network) and Machine Learning for High-Density WLANs. Proc. of National Conf. on Machine Learning (mar 2019)
35. Song, C., Park, Y., Golani, K., Kim, Y., Bhatt, K., Goswami, K.: Machine-learning based threat-aware system in software defined networks. ICCCN 2017 (sep 2017). https://doi.org/10.1109/ICCCN.2017.8038436
36. Sultana, N., Chilamkurti, N., Peng, W., Alhadad, R.: Survey on SDN based network intrusion detection system using machine learning approaches. Peer-to-Peer Networking and Applications 2018 12:2 **12**(2), 493–501 (jan 2018). https://doi.org/10.1007/S12083-017-0630-0
37. Tang, F., Li, L., Barolli, L., Tang, C.: An efficient sampling and classification approach for flow detection in SDN-based big data centers. AINA 2017 pp. 1106–1115 (may 2017). https://doi.org/10.1109/AINA.2017.125
38. Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K.L.A., Elkhatib, Y., Hussain, A., Al-Fuqaha, A.: Unsupervised Machine Learning for Networking: Techniques, Applications and Research Challenges. IEEE Access **7**, 65579–65615 (2019). https://doi.org/10.1109/ACCESS.2019.2916648
39. Wang, P., Lin, S.C., Luo, M.: A framework for QoS-aware traffic classification using semi-supervised machine learning in SDNs. SCC 2016 pp. 760–765 (aug 2016). https://doi.org/10.1109/SCC.2016.133
40. Wohlin, C.: Guidelines for snowballing in systematic literature studies and a replication in software engineering. EASE 2014 (2014). https://doi.org/10.1145/2601248.2601268
41. Xie, J., Richard Yu, F., Huang, T., Xie, R., Liu, J., Wang, C., Liu, Y.: A survey of machine learning techniques applied to software defined networking

(SDN): Research issues and challenges. COMST 2018 **21**(1), 393–430 (jan 2019). https://doi.org/10.1109/COMST.2018.2866942
42. Yu, C., Lan, J., Guo, Z., Hu, Y.: DROM: Optimizing the Routing in Software-Defined Networks with Deep Reinforcement Learning. IEEE Access **6**, 64533–64539 (2018). https://doi.org/10.1109/ACCESS.2018.2877686