AutoQML: Automated Quantum Machine Learning for Wi-Fi Integrated Sensing and Communications

Koike-Akino, Toshiaki; Wang, Pu; Wang, Ye

TR2022-068 July 20, 2022

Abstract

Commercial Wi-Fi devices can be used for integrated sensing and communications (ISAC) to jointly exchange data and monitor indoor environment. In this paper, we investigate a proof-of-concept approach using automated quantum machine learning (AutoQML) framework called AutoAnsatz to recognize human gesture. We address how to efficiently design quantum circuits to configure quantum neural networks (QNN). The effectiveness of AutoQML is validated by an in-house experiment for human pose recognition, achieving state-of-theart performance greater than 80% accuracy for a limited data size with a significantly small number of trainable parameters. Index Terms—Integrated sensing and communication (ISAC), Wi-Fi sensing, human monitoring, quantum machine learning.

IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM) 2022

AutoQML: Automated Quantum Machine Learning for Wi-Fi Integrated Sensing and Communications

Toshiaki Koike-Akino, Pu Wang, Ye Wang
Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA
Email: {koike, pwang, yewang}@merl.com

***Abstract*—Commercial Wi-Fi devices can be used for integrated sensing and communications (ISAC) to jointly exchange data and monitor indoor environment. In this paper, we investigate a proof-of-concept approach using automated quantum machine learning (AutoQML) framework called *AutoAnsatz* to recognize human gesture. We address how to efficiently design quantum circuits to configure quantum neural networks (QNN). The effectiveness of AutoQML is validated by an in-house experiment for human pose recognition, achieving state-of-the-art performance greater than 80% accuracy for a limited data size with a significantly small number of trainable parameters.**

***Index Terms*—Integrated sensing and communication (ISAC), Wi-Fi sensing, human monitoring, quantum machine learning.**

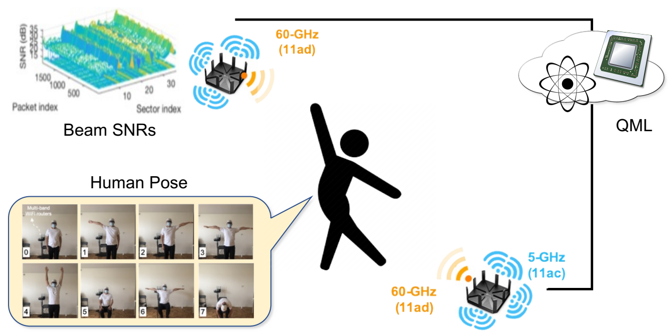


Fig. 1. Wi-Fi sensing for human pose recognition empowered by QML.

I. Introduction

A new paradigm called "integrated sensing and communications (ISAC)" has emerged as a key technology for future wireless systems [1]. In fact, many scenarios in the fifth generation and beyond (B5G), such as autonomous vehicles and extended reality, requires both high-performance sensing and data access. In addition, millimeter wave (mmWave) and massive multiple-input multiple-output (MIMO) technologies used in B5G can achieve high resolution in both time and angular domain, bringing ISAC a viable concept. In particular, Wi-Fi-based human monitoring has received much attention due to the decreasing cost and less privacy concerns compared with camera-based approaches. Modern deep neural networks (DNNs) have made Wi-Fi-band signals useful for user identification, emotion sensing, and skeleton tracking [2]–[25].

In our recent work [26], we introduced an emerging framework "quantum machine learning (QML)" [27]–[40] into ISAC applications for the first time in literature, envisioning future era of *quantum supremacy* [41, 42]. QML is considered as a key driver in the sixth generation (6G) applications [43], while there are few research yet to tackle practical problems. Quantum computers have the potential to realize computationally efficient signal processing compared to traditional digital computers by exploiting quantum mechanism, e.g., superposition and entanglement, in terms of not only execution time but also energy consumption. In the past few years, several vendors have successfully manufactured commercial quantum processing units (QPUs). For instance, IBM released 127-qubit QPUs in 2021, and plans to produce 1121-qubit QPUs by 2023. It is no longer far future when noisy intermediate-scale quantum (NISQ) devices [44] will be widely used for various real applications. Recently, hybrid quantum-classical algorithms based on the *variational* principle [45]–[48] were proposed to deal with quantum noise for NISQ devices.

In this paper, we further extend our previous work towards automated QML (AutoQML) [35, 36], which we refer to as *AutoAnsatz*, to facilitate quantum neural network (QNN) architecture design. Most QNN models [40] are based on particular quantum circuit templates known as *ansatz*, which still requires careful tuning of hyperparameters such as the number of qubits, the number of entangling layers, etc. Finding suitable ansatz and hyperparameters generally involves a considerable amount of manual trial-and-error. The AutoQML [35, 36] employs automated machine learning (AutoML) framework such as Bayesian optimization [49] to automate QML design. We experimentally validate the benefit of AutoQML in human pose recognition task using commercial-off-the-shelf (COTS) Wi-Fi devices. The contributions of this paper are four-fold as described below.

- This paper is the very first proof-of-concept study introducing AutoQML for Wi-Fi ISAC applications.
- We verify its feasibility for the human pose recognition application with COTS Wi-Fi devices.
- We consider a diverse variety of ansatz to design QNN.
- We validate that the optimized QNN via AutoAnsatz achieves high accuracy comparable to state-of-the-art DNN while the QNN is configured with a significantly smaller number of trainable parameters.

II. Quantum-Empowered Wi-Fi ISAC Systems

Fig. 1 shows the Wi-Fi human monitoring systems empowered by QML. During data communications, we collect beam scanning measurements associated with a class of human gestures as a fingerprinting data to learn QNN models.

A. Human Pose Recognition Experiments

Our experimental configuration uses two Wi-Fi stations: one station in front of a subject and another station behind the subject. Both stations are placed on a stand of a height of 1.2 meters with a distance of approximately 2 meters. The subject is asked to perform a total of 8 poses including distinct gestures like 'sit', 'stand with left arm lifted', etc. For each pose, we recorded 7 independent measurement sessions with different time duration and with sufficient time separation between consecutive two sessions. We use the measurements in the first four sessions as a training data and those in the last three sessions as a testing data. The total number of measurement samples is $42{,}915$ and $1{,}040$ in the training and testing, respectively.

B. COTS Wi-Fi Testbed: mmWave Beam SNR

As super-grained mmWave channel state information is not generally accessible from COTS devices without additional overhead, we use *mid-grained* Wi-Fi measurements in the beam angle domain—beam signal-to-noise ratios (SNRs)—generated from the beam training (a.k.a. beam alignment) phase. For each probing beampattern (a.k.a. beam sectors), beam SNR is collected by 802.11ad devices as a measure of beam quality. Such beam training is periodically carried out to adapt beam sectors to environmental changes. Accessing raw mmWave beam SNR measurements from COTS devices is enabled via an open-source software [50].

We use 802.11ad-compliant TP-Link Talon AD7200 routers to collect beam SNRs at 60 GHz. This router supports a single stream communication using analog beamforming over a 32-element planar array. From one beam training, one Wi-Fi station can collect 36 beam SNRs across discrete transmitting beampatterns. The measured beam SNRs are sent to a workstation via Ethernet cables to train a machine learning (ML) model over a cloud or on premise. The experimental system is deployed in a standard indoor room setting. Further details of the experiments can be found in our previous work [25].

C. Quantum Machine Learning (QML)

In [26], we introduced QML framework to the Wi-Fi sensing systems, considering the rapid growth of quantum technology. A number of modern DNN methods have been already migrated into quantum domain, e.g., convolutional layer [27], autoencoder [28], graph neural network [32], and generative adversarial network [30, 31]. Interestingly, the number of QML articles has been exponentially increasing at the same rate of DNN articles, doubling every year but just 6 years behind. It suggests that QML will be potentially used in numerous communities in a couple of years. More importantly, QML is more suited for Wi-Fi sensing because a cloud quantum server such as IBM QX and Amazon braket is readily accessible.

In analogy with DNN, it was proved that QNN holds the universal approximation property [51]. Accordingly, increasing the number of qubits may enjoy state-of-the-art DNN performance. In addition, quantum circuits are analytically differentiable [52], enabling stochastic gradient optimization of QNN. Nevertheless, QNN often suffers from a vanishing gradient issue called the barren plateau [53]. To mitigate the issue, a simplified 2-design (S2D) ansatz [34] was proposed to realize shallow entanglers for arbitrary unitary approximation. It is highly expected that quantum computers would offer breakthroughs in a wide range of fields. As classical deep learning has become extremely energy intensive [54], it is of importance to explore diverse computing modalities such as quantum computers for a future sustainable society.

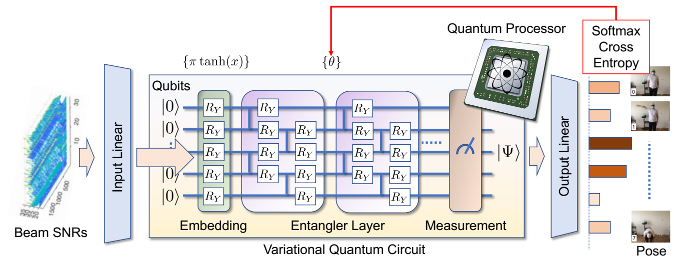


Fig. 2. Variational QNN for pose recognition.

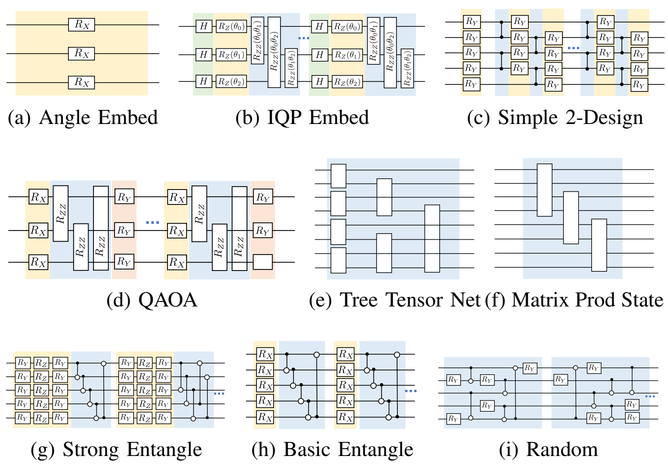
D. Quantum Neural Network (QNN) for Wi-Fi Sensing

Fig. 2 depicts the QNN model used for Wi-Fi sensing in our previous paper [26], employing S2D ansatz [34], which consists of Pauli-Y rotations and staggered controlled-Z entanglers. This ansatz is a simplified variant of a 2-design whose statistical properties are identical to ensemble random unitaries with respect to the Haar measure up to the first 2 moments. For an n-qubit variational quantum circuit, there are $2(n-1)L$ variational parameters $\{\theta\}$ over an L-layer S2D ansatz.

To feed 36-dimensional beam SNRs, an input linear layer is used to initialize the quantum state for rotation angles of Pauli-Y gates. The 8-class pose estimation is provided by quantum measurements in the Hamiltonian observable of Pauli-Z operations, followed by an output layer to align the dimension. The variational parameters as well as input/output layers are optimized by the adaptive momentum gradient method to minimize the softmax cross entropy loss. While QNN is not necessarily better than DNN in prediction accuracy, it can be computationally efficient by manipulating 2^n quantum states in parallel with a small number of quantum gates.

E. Automated QML for Ansatz/Hyperparameter Tuning

The best QNN hyperparameters (e.g., the qubit size n and layer size L) highly depend on datasets, and thus a considerable amount of manual effort is required to tune them in general. In addition, there are various types of quantum ansatz in literature to explore besides S2D. Fig. 3 illustrates some potential ansatz [40]: angle embedding; instantaneous quantum polynomial time (IQP) embedding [37]; quantum approximate optimization algorithm (QAOA) [45]; tensor network [33] including tree tensor network (TTN) and matrix product state (MPS); basic entangler; strongly entangling layers; random



(a) Angle Embed (b) IQP Embed (c) Simple 2-Design

(d) QAOA (e) Tree Tensor Net (f) Matrix Prod State

(g) Strong Entangle (h) Basic Entangle (i) Random

Fig. 3. Quantum ansatz variety for QNN architecture.

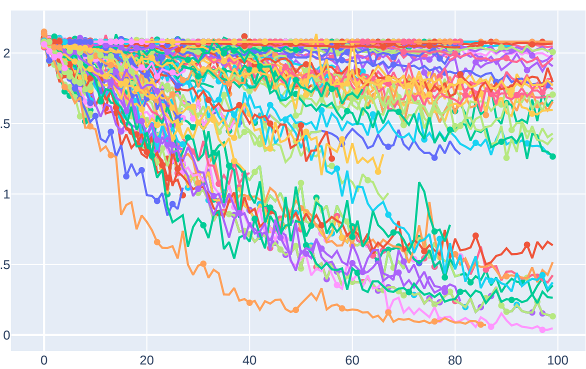


Fig. 4. Validation loss vs. learning epoch while AutoQML exploration with hyperband pruning (0.3% training data).

layers. In this paper, we propose *AutoAnsatz* in the context of AutoQML to optimize the QNN ansatz and hyperparameters. Specifically, we use Bayesian optimization based on tree-Parzen estimator and hyperband pruning [49]. It automatically explores the diverse set of quantum ansatz, qubit size, and layer size as well as learning rate, without the need of human effort in ansatz/hyperparameter tuning.

III. Performance Evaluation

A. AutoAnsatz Exploration Results

AutoQML explores 2,000 trials of hyperparameter tuning, where each model is trained with adaptive momentum (AdamW) gradient method over a maximum of 100 epochs. The Bayesian optimization uses categorical sampling of seven different ansatz and two different embedding methods as shown in Fig. 3. The number of qubits n and the number of entangler layers L are also sampled from a range of $[5, \ldots, 15]$ and $[1, \ldots, 5]$, respectively. In addition, the initial learning rate is optimized from a range of $(10^{-3}, 10^{-1})$, while the learning rate is adaptively decreased on plateau of training loss by a factor of 0.5 over a patience of 10 epochs.

Fig. 4 shows learning trajectories over training epoch, illustrating that some trials are efficiently pruned by the hyperband strategy due to hopeless or diverging trends. The pruning mechanism is important to accelerate automated ansatz search by early stopping useless trials. One can see that there are many trials having extremely slow convergence. The number of such pruned trials was about 89% of total trials.

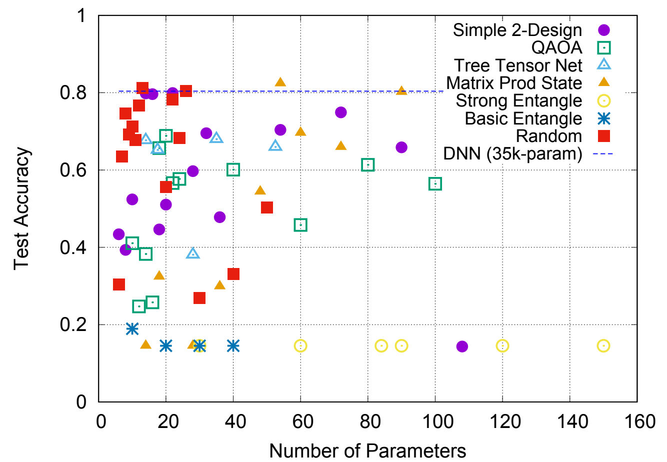


Fig. 5. Test accuracy vs. number of trainable parameters for various ansatz.

After AutoQML exploration, the best ansatz was chosen with 3-layer MPS for 10-qubit and angle embedding, having 54 variational parameters. Fig. 5 shows the test accuracy as a function of the total number of variational parameters for various models explored during the AutoQML. It shows the tradeoff between performance and complexity. The MPS ansatz tends to have better performance with more parameters, while most ansatz do not follow obvious trends. It is partly because we constrained the number of learning epoch up to 100, leading to poor underfitting performance for large-size QNN. The best MPS model outperforms the state-of-the-art DNN model which has 650-fold more trainable parameters (i.e., $35{,}000$). We can also observe that S2D and random layers ansatz offer a good tradeoff between accuracy and complexity, achieving DNN performance at $2{,}000$-fold fewer parameters.

B. AutoAnsatz Analysis

Fig. 6 shows the relation between training loss and individual hyperparameter. For ansatz search, we can see that strongly entangling and basic entangler layers perform poorly. For embedding, IQP was not optimized well, partly because of complicated high-order ansatz for differentiation. For learning rate, there is an obviously good choice at around 0.02. For the number of entangler layers, shallower QNNs tend to work better. The number of qubits above 8 (i.e., the number of pose classes) works equally well.

The loss contour is plotted in Fig. 7 showing a non-trivial landscape for the combination of two hyperparameters, indicating difficulty for manual tuning. Fig. 8 presents the functional analysis of variance (fANOVA) score [49] to assess the importance of hyperparameters. It is found that the choice of ansatz is the most influential hyperparameter, while the

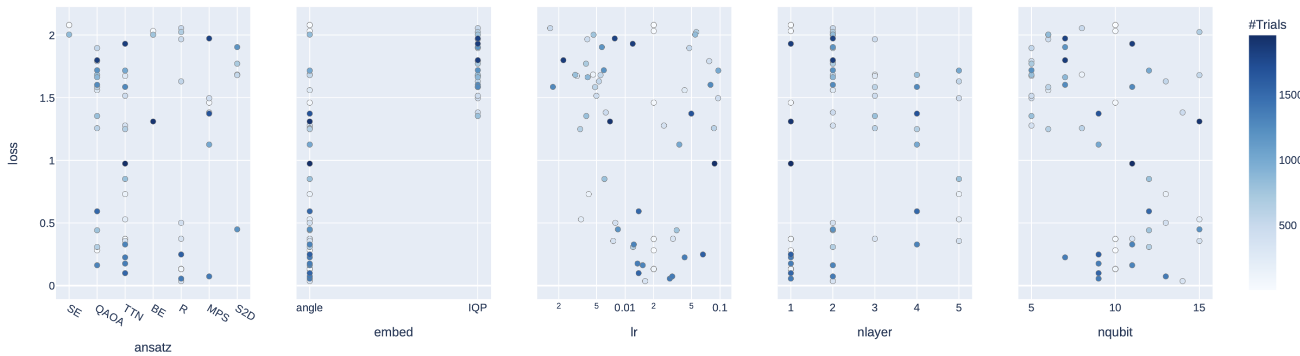


Fig. 6. Slice mapping of validation loss vs. single hyperparameter while AutoQML exploration.

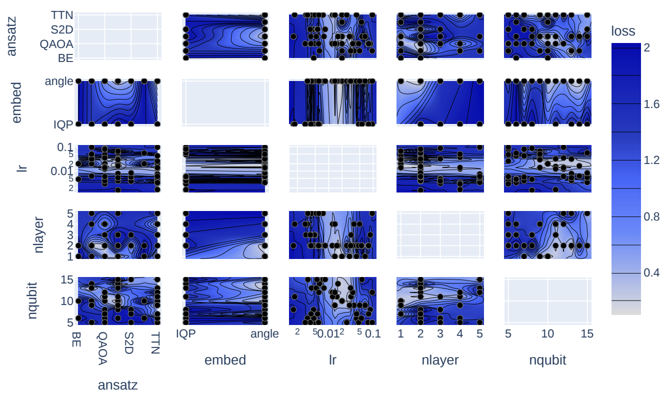


Fig. 7. Contour landscape of validation loss vs. two hyperparameters.

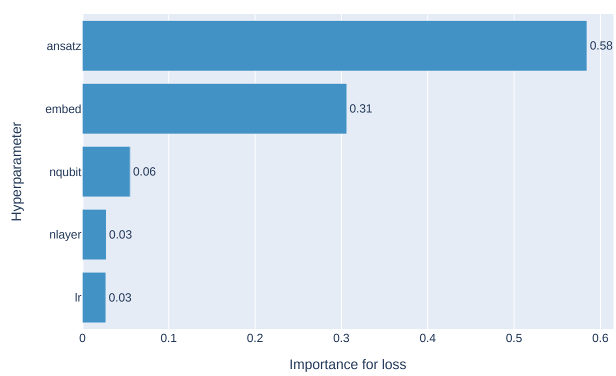


Fig. 8. fANOVA score on importance of QNN hyperparameters.

learning rate (which is usually most important for DNN design) is the lowest important one.

C. Comparison of Machine Learning (ML) Models

Finally, we compare different ML models including support vector machine (SVM), decision tree (DT), k-nearest neighbor (kNN), Gaussian naïve Bayes (GNB), random forest (RF), and extra tree ensemble methods (10 base models). For baseline DNN, we consider residual 4 hidden layers with 100 hidden nodes using Mish activation, with approximately 35k trainable parameters. For baseline QNN, we use 10-qubit 1-layer S2D ansatz [26], having 18 quantum variational parameters. We

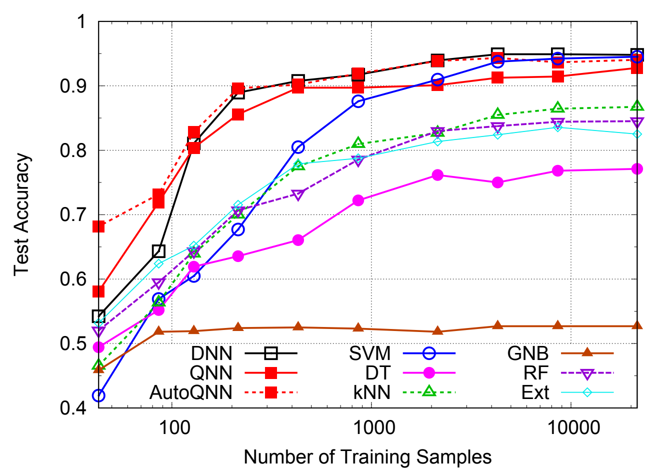


Fig. 9. Test accuracy vs. training samples for various ML methods.

use AdamW optimizer for a mini batch size of 100 over 100 epochs with a learning rate of 0.02 and weight decay of 10^{-4}.

Fig. 9 shows the test accuracy as a function of the number of labeled training samples. One can notice that the performance can exceed an accuracy of 90% for DNN, QNN, and SVM when a sufficient amount of labeled data is available. It is confirmed that a small-scale QNN designed by AutoQML can improve the baseline QNN model, achieving state-of-the-art performance comparable to a large-scale DNN.

IV. Conclusion

We investigated AutoQML to design ansatz of QNN for Wi-Fi sensing tasks. We demonstrated the benefit of AutoQML/AutoAnsatz through real-world experiments with an in-house Wi-Fi testbed. It was shown that a small-scale QNN can achieve state-of-the-art performance comparable to a large-scale DNN, in human pose recognition. Validation on real quantum processors will be provided in a future work. This is a very initial proof-of-concept study for quantum-ready Wi-Fi sensing and there remain many fascinating open issues.

Acknowledgment

We thank Jianyuan Yu (Virginia Tech.) for earlier data collection experiments.

REFERENCES

[1] A. Liu *et al.*, "A survey on fundamental limits of integrated sensing and communication," *IEEE Communications Surveys & Tutorials*, 2022.

[2] F. Adib *et al.*, "Capturing the human figure through a wall," *ACM Trans. Graph.*, vol. 34, no. 6, Oct. 2015.

[3] C.-Y. Hsu *et al.*, "Extracting gait velocity and stride length from surrounding radio signals," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 2116–2126.

[4] M. Zhao *et al.*, "Through-wall human pose estimation using radio signals," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2018, pp. 7356–7365.

[5] M. Zhao *et al.*, "RF-based 3D skeletons," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, 2018, pp. 267–281.

[6] P. Zhao *et al.*, "mID: Tracking and identifying people with millimeter wave radar," in *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, May 2019, pp. 33–40.

[7] A. Singh *et al.*, "RadHAR: Human activity recognition from point clouds generated through a millimeter-wave radar," in *Proceedings of the 3rd ACM Workshop on Millimeter-Wave Networks and Sensing Systems*, 2019, pp. 51–56.

[8] X. Lu *et al.*, "Robust occupancy inference with commodity WiFi," in *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Oct 2016, pp. 1–8.

[9] Y. Zeng *et al.*, "WiWho: WiFi-based person identification in smart spaces," in *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, April 2016, pp. 1–12.

[10] D. Wu *et al.*, "Device-free WiFi human sensing: From pattern-based to model-based approaches," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 91–97, Oct 2017.

[11] H. Zou *et al.*, "WiFi-based human identification via convex tensor shapelet learning," in *Proceedings of the 2018 AAAI Conference on Artificial Intelligence*, 2018, pp. 1711–1718.

[12] J. Yang *et al.*, "Device-free occupant activity sensing using WiFi-enabled IoT devices for smart homes," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3991–4002, Oct 2018.

[13] Y. Gu *et al.*, "Emosense: Data-driven emotion sensing via off-the-shelf WiFi devices," in *2018 IEEE International Conference on Communications (ICC)*, May 2018, pp. 1–6.

[14] ——, "Besense: Leveraging WiFi channel data and computational intelligence for behavior analysis," *IEEE Computational Intelligence Magazine*, vol. 14, no. 4, pp. 31–41, Nov 2019.

[15] Z. Wang *et al.*, "A survey on human behavior recognition using channel state information," *IEEE Access*, vol. 7, pp. 155 986–156 024, 2019.

[16] F. Wang *et al.*, "Person-in-WiFi: Fine-grained person perception using WiFi," in *Proceedings of (ICCV) International Conference on Computer Vision*, July 2019.

[17] F. Wang *et al.*, "Joint activity recognition and indoor localization with WiFi fingerprints," *IEEE Access*, vol. 7, pp. 80 058–80 068, 2019.

[18] Y. Ma *et al.*, "WiFi sensing with channel state information: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1–36, 2019.

[19] L. Zhang *et al.*, "WiVi: A ubiquitous violence detection system with commercial WiFi devices," *IEEE Access*, vol. 8, pp. 6662–6672, 2020.

[20] W. Jiang *et al.*, "Towards 3D human pose construction using WiFi," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1–14.

[21] M. Pajovic *et al.*, "Fingerprinting-based indoor localization with commercial MMWave WiFi–Part I: RSS and beam indices," in *2019 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2019.

[22] P. Wang *et al.*, "Fingerprinting-based indoor localization with commercial MMWave WiFi–Part II: Spatial beam SNRs," in *2019 IEEE Global Communications Conference (GLOBECOM)*, Dec 2019.

[23] T. Koike-Akino *et al.*, "Fingerprinting-based indoor localization with commercial MMWave WiFi: A deep learning approach," *IEEE Access*, vol. 8, pp. 84 879–84 892, 2020.

[24] P. Wang *et al.*, "Fingerprinting-based indoor localization with commercial mmwave WiFi: NLOS propagation," in *2020 IEEE Global Communications Conference (GLOBECOM)*, 2020, pp. 1–6.

[25] J. Yu *et al.*, "Human pose and seat occupancy classification with commercial MMWave WiFi," in *2020 IEEE Globecom Workshops*, 2020, pp. 1–6.

[26] T. Koike-Akino *et al.*, "Quantum transfer learning for Wi-Fi sensing," in *2022 IEEE International Conference on Communications (ICC)*. IEEE, 2022, pp. 1–6.

[27] M. Henderson *et al.*, "Quanvolutional neural networks: powering image recognition with quantum circuits," *Quantum Machine Intelligence*, vol. 2, no. 1, pp. 1–9, 2020.

[28] J. Romero *et al.*, "Quantum autoencoders for efficient compression of quantum data," *Quantum Science and Technology*, vol. 2, no. 4, p. 045001, 2017.

[29] P. Rebentrost *et al.*, "Quantum support vector machine for big data classification," *Physical review letters*, vol. 113, no. 13, p. 130503, 2014.

[30] S. Lloyd and C. Weedbrook, "Quantum generative adversarial learning," *Physical review letters*, vol. 121, no. 4, p. 040502, 2018.

[31] P.-L. Dallaire-Demers and N. Killoran, "Quantum generative adversarial networks," *Physical Review A*, vol. 98, no. 1, p. 012324, 2018.

[32] G. Verdon *et al.*, "Quantum graph neural networks," *arXiv preprint arXiv:1909.12264*, 2019.

[33] W. Huggins *et al.*, "Towards quantum machine learning with tensor networks," *Quantum Science and technology*, vol. 4, no. 2, p. 024001, 2019.

[34] M. Cerezo *et al.*, "Cost function dependent barren plateaus in shallow parametrized quantum circuits," *Nature communications*, vol. 12, no. 1, pp. 1–12, 2021.

[35] H. Wang *et al.*, "QuantumNAS: Noise-adaptive search for robust quantum circuits," *arXiv preprint arXiv:2107.10845*, 2021.

[36] R. B. Gómez *et al.*, "Towards autoqml: A cloud-based automated circuit architecture search framework," *arXiv preprint arXiv:2202.08024*, 2022.

[37] V. Havlíček *et al.*, "Supervised learning with quantum-enhanced feature spaces," *Nature*, vol. 567, no. 7747, pp. 209–212, 2019.

[38] M. Schuld *et al.*, "Circuit-centric quantum classifiers," *Physical Review A*, vol. 101, no. 3, p. 032308, 2020.

[39] E. Farhi and H. Neven, "Classification with quantum neural networks on near term processors," *arXiv preprint arXiv:1802.06002*, 2018.

[40] V. Bergholm *et al.*, "Pennylane: Automatic differentiation of hybrid quantum-classical computations," *arXiv preprint arXiv:1811.04968*, 2018.

[41] F. Arute *et al.*, "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, no. 7779, pp. 505–510, 2019.

[42] H.-S. Zhong *et al.*, "Quantum computational advantage using photons," *Science*, vol. 370, no. 6523, pp. 1460–1463, 2020.

[43] S. J. Nawaz *et al.*, "Quantum machine learning for 6G communication networks: State-of-the-art and vision for the future," *IEEE Access*, vol. 7, pp. 46 317–46 350, 2019.

[44] K. Bharti *et al.*, "Noisy intermediate-scale quantum (NISQ) algorithms," *arXiv preprint arXiv:2101.08448*, 2021.

[45] E. Farhi *et al.*, "A quantum approximate optimization algorithm," *arXiv preprint arXiv:1411.4028*, 2014.

[46] E. Farhi and A. W. Harrow, "Quantum supremacy through the quantum approximate optimization algorithm," *arXiv preprint arXiv:1602.07674*, 2016.

[47] E. Anschuetz *et al.*, "Variational quantum factoring," in *International Workshop on Quantum Technology and Optimization Problems*. Springer, 2019, pp. 74–85.

[48] A. Kandala *et al.*, "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets," *Nature*, vol. 549, no. 7671, pp. 242–246, 2017.

[49] T. Akiba *et al.*, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2623–2631.

[50] D. Steinmetzer *et al.*, "Talon tools: The framework for practical IEEE 802.11ad research," in *Available: https://seemoo.de/talon-tools/*, 2018.

[51] A. Pérez-Salinas *et al.*, "Data re-uploading for a universal quantum classifier," *Quantum*, vol. 4, p. 226, 2020.

[52] M. Schuld *et al.*, "Evaluating analytic gradients on quantum hardware," *Physical Review A*, vol. 99, no. 3, p. 032331, 2019.

[53] J. R. McClean *et al.*, "Barren plateaus in quantum neural network training landscapes," *Nature communications*, vol. 9, no. 1, pp. 1–6, 2018.

[54] E. Strubell *et al.*, "Energy and policy considerations for deep learning in NLP," *arXiv preprint arXiv:1906.02243*, 2019.