ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-6874-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICASSP49660.2025.10890305

Projection Valued-based Quantum Machine Learning Adapting to Differential Privacy Algorithm for Word-level Lipreading

Hang Chen[†], Chang Wang[†], Jun Du[†*], Chao-Han Huck Yang[‡], Jun Qi[§]
[†]NERC-SLIP, University of Science and Technology of China (USTC), Hefei, China
[‡]NVIDIA Research, Taibei, Taiwan [§]Hong Kong Baptist University, Hong Kong, China
* ✉jundu@ustc.edu.cn

***Abstract*—Deep neural network (DNN)-based lipreading models have achieved excellent recognition accuracy but are currently facing challenges related to user privacy. To address this, we propose a novel hybrid quantum-classical neural network (HQCNN) for lipreading that balances superior performance with enhanced privacy protection. The HQCNN-based lipreading model features an innovative variational quantum circuit (VQC) back-end, which transforms the output of the DNN front-end into quantum representations and predicts the posterior probability of each word. Furthermore, we introduce projection-valued encoding (PVE) and projection-valued measurement (PVM), enabling the VQC to handle inputs and outputs of dimensions that scale exponentially with the number of qubits, thereby substantially increasing its expressive power. Additionally, we explore the privacy-preserving properties of the HQCNN-based lipreading model by integrating differentially private stochastic gradient descent (DP-SGD). Experiments conducted on the LRW dataset demonstrate the model's exceptional recognition accuracy and privacy-preserving capabilities.**

***Index Terms*—Quantum machine learning, visual speech recognition, quantum multi-classifier, differential privacy.**

I. Introduction

Lipreading aims to identify spoken words by analyzing only the visual cues of a person's mouth and face [1] and can be classified into two types: word-level and sentence-level based on the recognition target. This study focuses on word-level lipreading (hereafter called ”lipreading”). This dynamic and challenging field holds immense potential for various domains [2] and has experienced significant breakthroughs over the past decade [3], [4], largely attributable to the advancement of deep neural networks (DNNs) [5]. However, DNN-based lipreading models are currently facing a user privacy challenge that leads to limited training data availability.

To address this issue, we propose integrating quantum machine learning (QML) [6], [7] into lipreading. QML offers unique quantum advantages over classical machine learning, leveraging the properties of the quantum bits (qubits), such as superposition and entanglement. The superposition ability of qubits to exist in multiple states concurrently allows QML algorithms to explore numerous possibilities at once, ensuring inherent privacy preservation capabilities. Additionally, the entanglement of qubits enables faster correlation calculations than classical algorithms, making algorithms like variational quantum circuits (VQC) [8]–[10] particularly adept at extracting complex and non-linear features from high-dimensional data [11], [12]. Unfortunately, due to limitations in current quantum computers, complex QML models face significant challenges in terms of scalability, trainability, and large-scale simulation [13]. These challenges stem from the computational instability and limited operators of early-stage quantum computers, in contrast to classical computers.

Consequently, we propose a novel hybrid quantum-classical neural network (HQCNN) for lipreading, leveraging the complementary strengths of QML and classical DNN in privacy preservation and complex model training. Specifically, we design an innovative VQC back-end that includes projection-valued encoding (PVE) and projection-valued measurement (PVM) to transform classical data into quantum representation and predict the posterior probability of each word. Our PVE and PVM methods are developed from positive operator valued encoding (POVE) and positive operator valued measurement (POVM) [14], allowing VQC to handle input and output of 2^n dimensions, significantly enhancing the expressive power and surpassing the recognition accuracy of classical DNNs. Additionally, we explore the privacy-preserving properties of the HQCNN-based lipreading model by incorporating the differentially private stochastic gradient descent (DP-SGD) [15], demonstrating its excellent privacy protection capability. The contributions of this work are summarized as follows:

1) propose a novel HQCNN-based lipreading model, balancing outstanding performance and user privacy protection.
2) enhancing the representative capability of quantum circuits by employing PVE and PVM in encoding and measurement, resulting in a significant improvement in recognition accuracy.
3) integrating the DP-SGD algorithm into model training that verifies the HQCNN-based lipreading model demonstrates superior privacy-preserving properties.

The remainder of this paper is organized as follows. Section II provides a review of relevant work. In Section III, we present a detailed description of the proposed hybrid quantum-classical neural network. Section IV details the experimental assessments conducted to evaluate and analyze the system's

*corresponding author

performance. Finally, Section V concludes the paper with a summary of our findings.

II. Related Work

A. Lipreading

Traditional lipreading systems typically begin with extracting manually engineered features from a lip-centered video and then decoding the features into word-level text [16]. However, in the past decade, end-to-end lipreading systems have gained increasing attention, replacing separate hand-engineered feature extractors and decoders with a whole DNN. For example, [17] proposed a prominent model backbone, combining a modified ResNet34 with a 2-layer bidirectional long short-term memory network. Subsequent research has focused on enhancing the model backbone with advancements such as the spatiotemporal fusion module [18], squeeze-and-extract module [19], time shift module [20], hierarchical pyramidal convolution [21], innovative alternating spatiotemporal and spatial convolution [22] and EfficientNet [23]. Recently, [4] introduced a novel VSP-LLM model leveraging large language models (LLMs) to enhance lipreading performance.

B. Quantum Machine Learning

Quantum machine learning (QML) stands out as an exciting application of quantum computing, offering computational efficiency and resistance to noise. Recent research [24] underscores the superiority of QML over classical machine learning, with benefits including reduced memory needs, secure parameter encryption, and enhanced representation capabilities. Notable advancements have been made in developing QML techniques as alternatives to traditional methods, encompassing quantum k-nearest neighbors [25], quantum support vector machines [26], quantum clustering [27], and quantum neural networks (QNNs) [28]. QNNs have especially proven successful across diverse domains. For example, [29] utilized quantum convolutional networks for image recognition, [30] proposed a quantum backpropagation system incorporating fuzzy logic for speech recognition, and [31] harnessed quantum transfer learning for synthetic speech detection. Furthermore, QNNs are being explored for privacy data processing in fields such as finance [32], chemistry [33], and healthcare [34], [35].

C. Differential Privacy

Differential privacy (DP) [36] ensures privacy safeguards by limiting the reliance on individual data points in data analysis algorithms. Formally, a randomized mechanism denoted as M satisfies (ϵ, δ)-differential privacy if, for any two datasets D and D' that differ by a single data point, and any subset S in the output space, we have $P[M(D) \in S] \leq e^{\epsilon} P\left[M\left(D'\right) \in S\right] + \delta$. Given its formidable flexibility and versatility, differential privacy methods have found wide-ranging applications. [37] successfully applying differential privacy to federated learning for speech recognition, [38] explores the effects of differential privacy on the training phase of the audio model. Moreover, [15] combined DP with stochastic gradient descent (SGD), which limits the influence of individual samples on the model and adds noise to protect privacy. [39] applied DP-SGD to the speech classification task, where DP-SGD updates parameters via privatized gradient descent steps, ensuring privacy during model training.

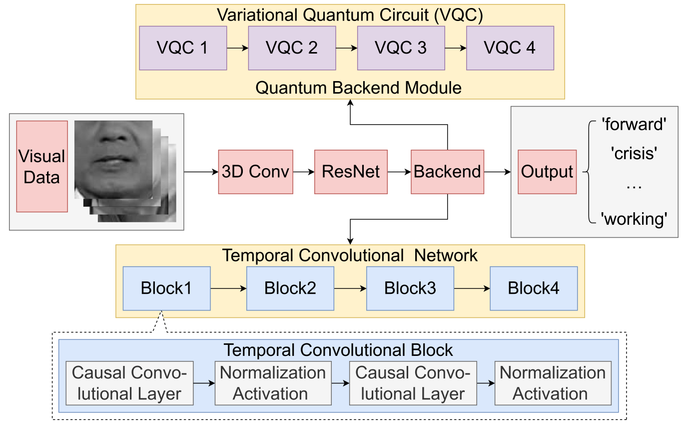


Fig. 1: Overall framework of the proposed hybrid quantum-classical neural network.

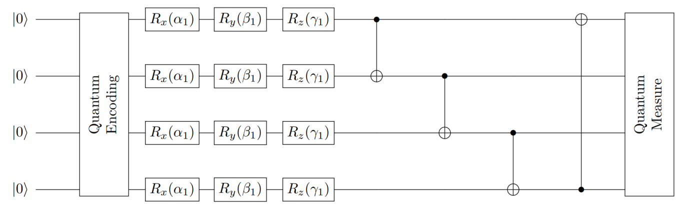


Fig. 2: Variational quantum circuit structure diagram.

III. Proposed Method

A. Overall Framework

Figure 1 illustrates the overall framework of the proposed hybrid quantum-classical neural network (HQCNN) for lipreading. The HQCNN-based model is divided into the front-end and back-end. In the front-end, a 3D convolutional layer followed by an 18-layer ResNet is employed, which is identical to the classical DNN-based lipreading model [40]. The key innovation lies in the back-end, where the HQCNN-based model incorporates a variational quantum circuit (VQC) as the back-end module, replacing the classical temporal convolutional network (TCN) [40].

The structure diagram of the VQC is shown in Fig. 2, where $|\alpha\rangle$ denotes the vector representation of a single qubit:

$$|\alpha\rangle = v_0|0\rangle + v_1|1\rangle \rightarrow \begin{bmatrix} v_0 \\ v_1 \end{bmatrix}. \tag{1}$$

Moreover, the VQC features a variety of quantum gates, each denoted by unitary matrices. The product of a matrix and the corresponding quantum state vector signifies the gate's operation on the quantum system. The unitary matrices associated with the Pauli rotation gates $R_X(\cdot), R_Y(\cdot), R_Z(\cdot)$ are shown below:

$$\begin{gathered} R_X\left(\theta\right) = \begin{bmatrix} \cos\frac{\theta}{2} & -i\sin\frac{\theta}{2} \\ -i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix} \\ R_Y\left(\theta\right) = \begin{bmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}, R_Z\left(\theta\right) = \begin{bmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{-i\frac{\theta}{2}}. \end{bmatrix} \end{gathered} \tag{2}$$

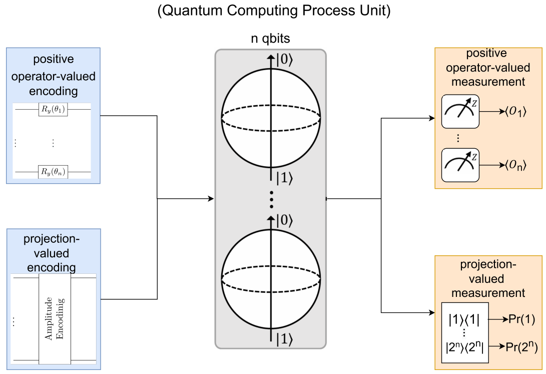


Fig. 3: Illustration of the encoding and decoding methods of the quantum circuit. On the left side, we represent two encoding methods: POVE and PVE. On the right side, we have two decoding methods: POVM and PVM.

These Pauli rotation gates separately correspond to a rotation around the X, Y, and Z axes of the Bloch sphere by θ radians. The controlled NOT gate (CNOT) acts on 2 qubits, and performs the NOT operation on the second qubit only when the first qubit is $|1\rangle$, and otherwise leaves it unchanged. With respect to the basis $|00\rangle$, $|01\rangle$, $|10\rangle$, $|11\rangle$, it is represented by the Hermitian unitary matrix:

$$\text{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{3}$$

It can be described as the gate that maps the basis states $|a, b\rangle \mapsto |a, a \oplus b\rangle$, where$\oplus$ is XOR.

B. Projection Valued-based Encoding and Measurements

The encoding and measurement methods of the quantum circuit are illustrated in Fig. 3. Currently, most quantum computers and QML frameworks utilize positive operator-valued encoding (POVE) to handle inputs and positive operator-valued measurements (POVM) to obtain outputs [14]. However, in an n-qubit system, the input size based on POVE and the output size based on POVM cannot exceed the number of qubits, i.e., n, thereby constraining the scalability of QNN. Consequently, we propose a novel QML model based on projection values, including the projection value encoding (PVE) and the projection value measurements (PVM). The PVE module directly initializes the quantum state, thereby elevating the input dimension that can be processed by the quantum circuit from n to 2^n. Meanwhile, the PVM module represents a specific form of POVM [41] and generates 2^n distinct outputs (referred to as observables) using the full basis of qubits. Through this approach, we significantly enhance the input-output capability of the quantum circuit, effectively improving its performance.

C. DP-SGD Algorithm

The specific procedure of the DP-SGD algorithm is depicted in Algorithm 1. We incorporate Gaussian noise into the computed training gradients in each mini-batch, ensuring

TABLE I: Accuracy of TCN and VQC back end modules. K_s and Q_n represent kernel size and qubits number respectively.

Block Number		1	2	3	4
TCN	$K_s = 3$	71.9	77.5	80.2	84.9
	$K_s = 5$	74.8	78.3	81.9	85.2
	$K_s = 7$	76.2	80.2	82.3	85.0
VQC	$Q_n = 3$	83.9	85.1	85.9	86.5
	$Q_n = 4$	84.9	85.8	86.3	86.7
	$Q_n = 5$	84.3	85.5	86.1	86.4

differential privacy. We effectively control the privacy budget by tuning the added noise's standard deviation. This enables us to compare the outcomes of the quantum backend module and the temporal convolutional network within the same privacy budget. Given that DP-SGD achieves differential privacy by perturbing the training gradients, we will also compute the average gradients of the quantum backend module and the temporal convolutional network in our subsequent experiments. This will enable us to discern the impact of noise of the same magnitude on different algorithms.

Algorithm 1 Differentially Private Stochastic Gradient Descent (DP-SGD)

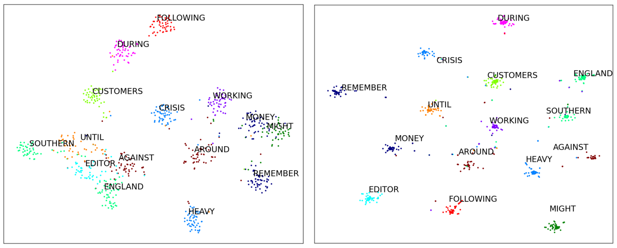
Input: Dataset $\mathcal{D}$, Noise parameter σ, Learning rate η
Output: Model parameters θ
Initialize θ randomly
for each epoch **do**
  Shuffle $\mathcal{D}$ randomly
  for each mini-batch (x, y) in $\mathcal{D}$ **do**
    Compute the gradient $\nabla_\theta \mathcal{L}(x, y, \theta)$
    Add Gaussian noise $N(0, \sigma^2)$ to $\nabla_\theta \mathcal{L}(x, y, \theta)$
    Model update: $\theta \leftarrow \theta - \eta \cdot \nabla_\theta \mathcal{L}(x, y, \theta)$

IV. Experiments

We conducted experiments with the Lip Reading in the Wild (LRW) dataset, consisting of video segments extracted from BBC television broadcasts. The LRW dataset includes 500 target words and multiple speakers, totaling around 500,000 segments. To ensure accurate alignment and representation of the mouth region for subsequent analysis, we performed the following preprocessing steps on each video sequence from the LRW dataset: 1) Face detection and alignment were applied.2) Each frame image was aligned with a reference average face shape. 3) A fixed 96×96 pixel region of interest (ROI) was cropped from the aligned face image, with the mouth region approximately centered within the cropped image.4)The cropped image was converted to grayscale. We also employ data augmentation of random cropping at 88×88 pixels and random horizontal flipping. Both augmentation methods are consistently applied to all frames within the sequence.

A. Analysis of VQC Back-end

We compared a quantum backend module, constructed using quantum circuits as building blocks, to a temporal convolutional network. For the temporal convolutional network, we



(a) TCN (b) VQC

Fig. 4: T-SNE visualization of TCN and VQC back-ends.

selected convolutional kernel sizes of 3, 5, and 7, while for the quantum circuits, we chose 3, 4, and 5 qubits. The experimental results are presented in Table I. Both methods' performance improves as the block size increases. The temporal convolutional network exhibits a higher growth rate, but the quantum backend module consistently outperforms it.

Subsequently, we randomly selected 15 words and employed the t-SNE method to visualize the corresponding embedding results generated by both approaches. The resulting visualization is depicted in Figure 4. From the figure, it can be observed that the embeddings produced by quantum neural networks exhibit a noticeably more compact arrangement compared to the scattered distribution of embeddings generated by conventional neural networks. Further analysis reveals that, in comparison to temporal convolutional networks, the quantum backend module exhibits superior capability in distinguishing words with similar pronunciation, such as "money" and "might", "ENGLAND" and "EDITOR".This phenomenon effectively demonstrates the superiority of QML.

B. *Analysis of PVE and PVM*

TABLE II: Accuracy (Acc) for different encoding and decoding methods of the quantum circuit.

Qbits Number	POVE & POVM	POVE & PVM	PVE & POVM	PVE & PVM
3	53.2	65.5	64.2	**79.1**
4	65.3	75.4	79.8	**84.3**
5	73.6	80.1	79.6	**85.1**
6	75.2	80.7	80.3	**84.7**
7	77.8	82.5	81.7	**85.0**
8	79.3	82.2	81.3	**84.5**

The experimental results of quantum circuits, employing different input-output methods, are presented in the table II. The observations from Table II indicate that combining PVE and PVM methods significantly enhances the expressive power of quantum circuits compared to the traditional POVE and POVM methods. This improvement is consistent across different numbers of qubits. The reason for this improvement is that in low-dimensional settings, a small number of qubits struggle to extract the necessary information for multi-classification tasks.

Notably, the results of the POVE & POVM method with 8 qubits are comparable to thosee of the PVE & PVM method with 3 qubits. This comparison is significant because it indicates that even with the same input and output dimensions

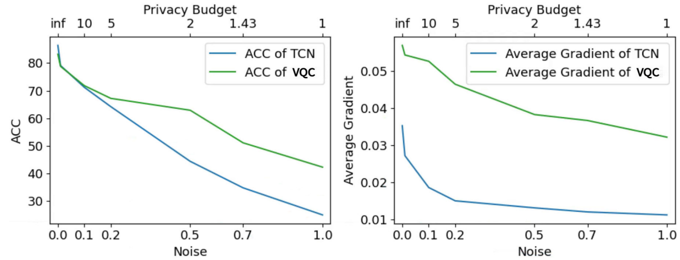


Fig. 5: Illustration of DP-SGD algorithm results. The left figure illustrates the recognition accuracy of TCN and VQC under various noise/privacy budget conditions, while the right figure represents the average gradient values of TCN and VQC.

(2^3 = 8), the dimension expansion achieved through PVE & PVM provides an advantage in performance.

C. *Analysis of DP-SGD*

The combined results of TCN and VQC with the DP-SGD algorithm are depicted in Figure 5. The left figure clearly shows that as the level of noise increases, both the quantum neural network and the conventional neural network experience a decline in performance. However, the quantum neural network exhibits a much smaller degree of decline compared to the conventional neural network. Considering that DP-SGD introduces noise to the training gradients, we calculate the average gradient for both networks, and the results are presented in the right figure. From the right figure, it is evident that the gradient of the VQC back-end significantly surpasses that of TCN. Consequently, it implies that VQC is less susceptible to the interference of identical noise than TCN. This observation explains why VQC exhibits superior resilience to noise when compared to TCN.

V. Conclusion

This paper proposes a novel HQCNN-based lipreading model featuring an innovative VQC back-end, which incorporates PVE and PVM methods to transform classical data into quantum representations and predict the posterior probability of each word. Experimental results demonstrate that the VQC back-end generates more discriminative representations than the baseline TCN back-end, significantly improving recognition accuracy. Additionally, the PVE and PVM methods achieve dimensional expansion, providing a clear advantage in the expressive power of the VQC compared to POVE and POVM methods with the same number of qubits. Finally, under the same level of noise introduced by DP-SGD, the gradient of the VQC back-end significantly outperforms that of the TCN, further demonstrating the superior privacy-preserving capabilities of the VQC back-end.

VI. Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under Grant 62171427, and in part by MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China.

REFERENCES

[1] Sarah L Taylor, Moshe Mahler, Barry-John Theobald, and Iain Matthews, "Dynamic units of visual speech," in *Proc. SIGGRAPH 2012*, 2012, pp. 275–284.

[2] Triantafyllos Afouras, Joon Son Chung, Andrew Senior, Oriol Vinyals, and Andrew Zisserman, "Deep audio-visual speech recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 12, pp. 8717–8727, 2018.

[3] Brendan Shillingford, Yannis Assael, Matthew W. Hoffman, Thomas Paine, Cían Hughes, Utsav Prabhu, Hank Liao, Hasim Sak, Kanishka Rao, Lorrayne Bennett, Marie Mulville, Misha Denil, Ben Coppin, Ben Laurie, Andrew Senior, and Nando de Freitas, "Large-scale visual speech recognition," in *Proc. Interspeech 2019*, 2019, pp. 4135–4139.

[4] Jeong Hun Yeo, Seunghee Han, Minsu Kim, and Yong Man Ro, "Where visual speech meets language: Vsp-llm framework for efficient and context-aware visual speech processing," *arXiv preprint arXiv:2402.15151*, 2024.

[5] Geoffrey E Hinton and Ruslan R Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[6] M Cerezo, Guillaume Verdon, Hsin-Yuan Huang, Lukasz Cincio, and Patrick J Coles, "Challenges and opportunities in quantum machine learning," *Nature Computational Science*, vol. 2, no. 9, pp. 567–576, 2022.

[7] Hsin-Yuan Huang et al., "Power of Data in Quantum Machine Learning," *Nature Communications*, vol. 12, no. 1, pp. 1–9, 2021.

[8] Samuel Yen-Chi Chen, Chao-Han Huck Yang, Jun Qi, Pin-Yu Chen, Xiaoli Ma, and Hsi-Sheng Goan, "Variational quantum circuits for deep reinforcement learning," *IEEE Access*, vol. 8, pp. 141007–141024, 2020.

[9] Marco Cerezo et al., "Variational Quantum Algorithms," *Nature Reviews Physics*, vol. 3, no. 9, pp. 625–644, 2021.

[10] Jun Qi, Chao-Han Yang, and Pin-Yu Chen, "QTN-VQC: An End-to-End Learning Framework for Quantum Neural Networks," *Physica Scripta*, vol. 99, no. 1, pp. 015111, 2023.

[11] Jun Qi, Chao-Han Huck Yang, Pin-Yu Chen, and Min-Hsiu Hsieh, "Theoretical error performance analysis for variational quantum circuit based functional regression," *npj Quantum Information*, vol. 9, no. 1, pp. 4, 2023.

[12] Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner, "The Power of Quantum Neural Networks," *Nature Computational Science*, vol. 1, no. 6, pp. 403–409, 2021.

[13] Arthur Pesah, Marco Cerezo, Samson Wang, Tyler Volkoff, Andrew T Sornborger, and Patrick J Coles, "Absence of barren plateaus in quantum convolutional neural networks," *Physical Review X*, vol. 11, no. 4, pp. 041011, 2021.

[14] Leonardo Banchi, Jason Pereira, and Stefano Pirandola, "Generalization in quantum machine learning: A quantum information standpoint," *PRX Quantum*, vol. 2, no. 4, pp. 040321, 2021.

[15] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang, "Deep learning with differential privacy," in *Proc. ACM SIGSAC CCS 2016*, 2016, pp. 308–318.

[16] Ziheng Zhou, Guoying Zhao, Xiaopeng Hong, et al., "A review of recent advances in visual speech decoding," *Image and Vision Computing*, vol. 32, no. 9, pp. 590–605, 2014.

[17] T. Stafylakis and G. Tzimiropoulos, "Combining residual networks with lstms for lipreading," in *Proc. Interspeech 2017*, 2017, pp. 3652–3656.

[18] Xingxuan Zhang, Feng Cheng, and Wang Shilin, "Spatio-temporal fusion based convolutional sequence learning for lip reading," in *Proc. ICCV 2019*, 2019, pp. 713–722.

[19] Dalu Feng, Shuang Yang, and Shiguang Shan, "An efficient software for building lip reading models without pains," in *Proc. ICMEW 2021*, 2021, pp. 1–2.

[20] Mingfeng Hao, Mutallip Mamut, Nurbiya Yadikar, et al., "How to use time information effectively? combining with time shift module for lipreading," in *Proc. ICASSP 2021*, 2021, pp. 7988–7992.

[21] H. Chen, J. Du, Y Hu, et al., "Automatic lip-reading with hierarchical pyramidal convolution and self-attention for image sequences with no word boundaries," in *Proc. INTERSPEECH 2021*, 2021, pp. 3001–3005.

[22] Dimitrios Tsourounis, Dimitris Kastaniotis, and Spiros Fotopoulos, "Lip reading by alternating between spatiotemporal and spatial convolutions," *Journal of Imaging*, vol. 7, no. 5, pp. 91, 2021.

[23] Alexandros Koumparoulis and Gerasimos Potamianos, "Accurate and resource-efficient lipreading with efficientnetv2 and transformers," in *Proc. ICASSP 2022*, 2022, pp. 8467–8471.

[24] Chao-Han Huck Yang, Bo Li, Yu Zhang, Nanxin Chen, Tara N Sainath, Sabato Marco Siniscalchi, and Chin-Hui Lee, "A quantum kernel learning approach to acoustic modeling for spoken command recognition," in *Proc. ICASSP 2023*, 2023, pp. 1–5.

[25] Yijie Dang, Nan Jiang, Hao Hu, Zhuoxiao Ji, and Wenyin Zhang, "Image classification based on quantum k-nearest-neighbor algorithm," *Quantum Information Processing*, vol. 17, pp. 1–18, 2018.

[26] Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd, "Quantum support vector machine for big data classification," *Physical Review Letters*, vol. 113, no. 13, pp. 130503, 2014.

[27] Thomas Maier, Mark Jarrell, Thomas Pruschke, and Matthias H Hettler, "Quantum cluster theories," *Reviews of Modern Physics*, vol. 77, no. 3, pp. 1027, 2005.

[28] Edward Farhi and Hartmut Neven, "Classification with quantum neural networks on near term processors," *arXiv preprint arXiv:1802.06002*, 2018.

[29] YaoChong Li, Ri-Gui Zhou, RuQing Xu, Jia Luo, and WenWen Hu, "A quantum deep convolutional neural network for image recognition," *Quantum Science and Technology*, vol. 5, no. 4, pp. 044003, 2020.

[30] Fei Li, Shengmei Zhao, and Baoyu Zheng, "Quantum neural network in speech recognition," in *Proc. ICSP 2002*, 2002, vol. 2, pp. 1267–1270.

[31] Ruoyu Wang, Jun Du, and Tian Gao, "Quantum transfer learning using the large-scale unsupervised pre-trained model wavlm-large for synthetic speech detection," in *Proc. ICASSP 2023*, 2023, pp. 1–5.

[32] Eric Paquet and Farzan Soleymani, "Quantumleap: Hybrid quantum neural network for financial predictions," *Expert Systems with Applications*, vol. 195, pp. 116583, 2022.

[33] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M Chow, and Jay M Gambetta, "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets," *Nature*, vol. 549, no. 7671, pp. 242–246, 2017.

[34] Michael Esposito, Glen Uehara, and Andreas Spanias, "Quantum machine learning for audio classification with applications to healthcare," in *Proc. IISA 2022*, 2022, pp. 1–4.

[35] Muhammad Junaid Umer, Javeria Amin, Muhammad Sharif, Muhammad Almas Anjum, Faisal Azam, and Jamal Hussain Shah, "An integrated framework for covid-19 classification based on classical and quantum transfer learning from a chest radiograph," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 20, pp. e6434, 2022.

[36] Abhishek Bhowmick, John Duchi, Julien Freudiger, Gaurav Kapoor, and Ryan Rogers, "Protection against reconstruction and its applications in private federated learning," *arXiv preprint arXiv:1812.00984*, 2018.

[37] Martin Pelikan, Sheikh Shams Azam, Vitaly Feldman, Jan Silovsky, Kunal Talwar, Tatiana Likhomanenko, et al., "Federated learning with differential privacy for end-to-end speech recognition," *arXiv preprint arXiv:2310.00098*, 2023.

[38] BN Suhas, Sarah Rajtmajer, and Saeed Abdullah, "Differential privacy enabled dementia classification: An exploration of the privacy-accuracy trade-off in speech signal data," in *Proc. INTERSPEECH 2023*, 2023, vol. 2023, pp. 346–350.

[39] Alexander Ziller, Dmitrii Usynin, Moritz Knolle, Kerstin Hammernik, Daniel Rueckert, and Georgios Kaissis, "Complex-valued deep learning with differential privacy," *arXiv preprint arXiv:2110.03478*, 2021.

[40] Brais Martinez, Pingchuan Ma, Stavros Petridis, and Maja Pantic, "Lipreading using temporal convolutional networks," in *Proc. ICASSP 2020*, 2020, pp. 6319–6323.

[41] Kurt Jacobs, *Quantum measurement theory and its applications*, Cambridge University Press, 2014.