Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)

Performance Analysis of Hybrid Quantum-Classical Convolutional Neural Networks for Audio Classification

Submitted in partial fulfillment of the requirement
of the degree of

Bachelor of Technology in

Department of Computer Science and Engineering (Data Science)

By

Bhuvi Ghosh	**60009210191**
Vishma Adeshra	**60009210201**
Yash Thakar	**60009210205**

Under the guidance of

Dr. Kriti Srivastava
Head of the Department

A.Y. 2023 – 2024

Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)

CERTIFICATE

This is to certify that the project entitled, **"Performance Analysis of Hybrid Quantum-Classical Convolutional Neural Networks for Audio Data "** is a bonafide work of **"Bhuvi Ghosh" (60009210191), "Vishma Adeshra" (60009210200) and "Yash Thakar" (6000921205)** submitted in the partial fulfillment of the requirement for the award of the Bachelor of Technology in Computer Science and Engineering (Data Science).

Dr. Kriti Srivastava
Guide

Dr Kriti Srivastava
Head of the Department

Dr Hari Vasudevan
Principal

Place: Mumbai
Date:

Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)

DECLARATION

We declare that this written submission represents our ideas in our own words and where others ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that, we have adhered to all the principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources, which have thus not been properly cited or from whom proper permission has not been taken, when needed.

Bhuvi Ghosh (60009210191)

Vishma Adeshra (60009210200)

Yash Thakar (60009210205)

Place: Mumbai

Date:

Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)

APPROVAL SHEET

Project entitled, **"Performance Analysis of Hybrid Quantum-Classical Convolutional Neural Networks for Audio Data"**, submitted by **"Bhuvi Ghosh" (60009210191), "Vishma Adeshra" (60009210200) and "Yash Thakar" (6000921205)** is approved for the award of the Bachelor of Technology in Computer Science and Engineering (Data Science).

Signature of Internal Examiner **Signature of External Examiner**

Place: Mumbai

Date:

Acknowledgement

We want to convey our heartfelt appreciation to Dr. Kriti Srivastava, our dedicated project supervisor, for her invaluable guidance, unwavering support, and insightful feedback throughout the development of the project. Dr. Srivastava's expertise has been a guiding force, shaping the project and ensuring its successful execution.

We would like to extend our sincere appreciation to Dr. Hari Vasudevan, the esteemed Principal of our institution, for his unwavering support and encouragement throughout the duration of our project. Dr. Hari Vasudevan's visionary leadership has created an academic environment that fosters innovation and research, providing us with the opportunity to undertake and excel in our endeavors.

Our gratitude also extends to the Computer Science and Engineering (Data Science) Department for providing essential resources and creating an environment conducive to our research and development activities. The department's facilities have played a crucial role in the seamless progression and accomplishment of our project goals.

Special thanks are due to the exceptional members of our project team, whose collective efforts and diverse skills have been fundamental in bringing this innovative concept to fruition. The commitment to excellence and effective collaboration demonstrated by each team member has been instrumental in the overall success of the project.

Abstract

Audio signals being high-dimensional and complex pose challenges for classical machine learning techniques in terms of computation and generalization on real-world data. This project evaluates the use of hybrid quantum-classical convolutional neural networks (QCNNs) that leverage quantum effects like superposition and entanglement for audio classification using mel-spectrograms obtained from audio data. Evaluated on both small-sized and large-sized datasets, the proposed hybrid QCNN model gave comparable training accuracy with classical CNN on the smaller dataset but outperformed classical CNN on test accuracy (95.04% vs 92.88%) for a larger birdsong dataset and reduced overfitting, thus highlighting the potential advantages of QCNNs for high-dimensional audio data. The hybrid QCNN exhibited higher cross-entropy loss in case of the small sized dataset which was further significantly reduced when evaluated on the large sized Birdsong data.

Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)

Table of Contents

List Of Tables

List Of Figures

List Of Abbreviations

QCNN	Quantum Convolutional Neural Network
QSVC	Quantum Support Vector Classifier
QML	Quantum Machine Learning
MFCC	Mel Frequency Cepstral Coefficients
CNN	Convolutional Neural Network
QNN	Quantum Neural Network

1. INTRODUCTION

Background:

Audio signals are rich in information, comprising various attributes such as frequency content, amplitude variations, and temporal patterns. Analyzing and classifying audio data pose significant challenges due to its complex and multi-dimensional nature. Traditional machine learning techniques struggle to effectively process such data due to its high complexity. One popular approach to tackle audio classification tasks is by using convolutional neural networks (CNNs) on spectrogram representations of audio signals. Mel-spectrograms provide a visual encoding of the frequency content of audio signals over time, allowing CNNs to identify intricate spatial patterns and features. Through techniques like mel-scale filtering, log transformation, and normalization, mel-spectrograms capture perceptually relevant features in a low-dimensional, time-frequency representation. This enables CNNs to effectively differentiate between various audio classes such as voice data, music genres, environmental sounds, etc.

Motivation:

Despite the success of CNNs in audio classification, the growing volume, variations, and dimensionality of audio data present challenges for classical computational methods. Classical computers face limitations in processing power, memory requirements, and classification accuracy when dealing with such high-dimensional data. This motivates exploration into the potential of quantum computing and quantum neural network (QNN) architectures for audio classification tasks. Quantum computers leverage principles of quantum mechanics like superposition and entanglement to perform computations exponentially faster than classical computers. Quantum convolutional neural networks (QCNNs) combine quantum computing with deep learning techniques, making them suitable for handling high-dimensional data like audio spectrograms. QCNNs exploit quantum parallelism and other quantum effects to potentially offer significant computational advantages over classical CNNs. This could lead to more efficient processing of high-dimensional audio data and improved classification accuracy across different audio classes.

2. LITERATURE SURVEY

A. Classical Machine Learning

[1] proposed a 6-layered CNN architecture for environmental sound classification trained on mel-spectrograms. They employed the MixUp data augmentation technique to generate new samples. On a 50-class environmental sound dataset their mel-spectrogram CNN achieved 81.75% accuracy, outperforming models like LeNet-5 (50.5%), AlexNet (57.15%), ResNet-50 (68.5%), and VGG11 (79.25%).

For music genre classification, [2] proposed a sequential deep learning model (ANN) operating directly on mel-spectrogram image inputs from the GTZAN dataset Their model attained 94% accuracy on the test set [3] explored classifying different types of baby cries using transfer learning, SVMs and ensemble techniques on the Baby Chillanto database with audio samples across 5 classes. They trained a baseline CNN model on audio spectrograms, a pre-trained ResNet50 model via transfer learning and an SVM model. By ensembling the predictions from the CNN and SVM, their combined model achieved a maximum accuracy of 91.10%, outperforming the individual models.

Kinha et al. [5] proposed novel architectures for heart sound classification from spectrograms on the PhysioNet dataset. Their ResNet18 model with Shuffle Attention layers, designed to refine learned features achieved superior performance with 91.27% accuracy, 94.75% recall, and 92.51% F1-score, outperforming the ResNeXt50 variant and traditional methods. Bansal et al. [4] developed a CNN architecture operating on MFCC features extracted from cough audio clips, achieving 70.58% accuracy, 80.95% recall, and 69.59% F1-score for classifying potential COVID-19 coughs on a dataset of 501 samples. Their spectrogram-based approach using transfer learning with a VGG16 model yielded less promising results, highlighting the importance of feature representation. A gated convolutional neural network for audio classification was proposed in [6] [7] used a BLSTM network on acoustic stages to improve audio classification. Audio Swin Transformer [8] which introduces a shifted window scheme for efficient self-attention computation and Conformer models [9] that combine convolutions and transformers have also achieved state-of-the-art performance for speech recognition, speaker verification, acoustic scene classification, and audio captioning.

B. Quantum Machine Learning

In the area of Quantum Machine Learning [10] proposed a Multiscale Entanglement Renormalization Ansatz (MERA) inspired ternary unitary circuit feature extraction method based on QCNN for binary image classification. The testing done on a breast cancer image dataset showed that their QCNN model achieved a recognition rate of 90-93% compared to 83-85% for the classical model

[11] highlights challenges such as the lack of real-time quantum computers, improper training algorithms, and the non-linear nature of neural networks. The paper implemented a QCNN model, finding that QCNNs can increase the computational speed when run on quantum computers with better performance metrics compared to classical methods for image recognition and object detection.

In the domain of remote sensing, [12] introduced a hybrid quantum-classical convolutional neural network (QC-CNN) for image classification using Earth observation (EO) data. Their approach involved using quantum neural networks to extract high-level critical features and used amplitude encoding to reduce quantum bit resources. On various EO benchmarks, the QC-CNN achieved better performance and higher generalizability than its classical counterpart, verifying its validity for EO data classification.

Recognizing the dimensionality challenges faced by classical CNNs, [13] proposed a QCNN model with a quantum encoding circuit that can encode classical data into quantum data, allowing the QCNN to run on quantum computers. They introduced quantum state amplitude encoding to represent high-dimensional classical data as quantum states, significantly reducing the dimensionality of the processed data [14] proposes an effective model called Adaptive Spotted Hyena Optimizer-based Deep Quantum Neural Network (ASHO-based Deep QNN) for automated classification of laryngeal cancer from endoscopic images. Classification is done with a Deep Quantum Neural Network where the weights are optimized using the novel ASHO algorithm, which combines the adaptive concept with the Spotted Hyena Optimizer. It achieved a maximum accuracy of 0.948, sensitivity of 0.952, and specificity of 0.924 for laryngeal cancer detection.

In [15] the authors propose a Pegasos Quantum Support Vector Classifier (QSVC) for classifying cardiovascular diseases, which outperformed classical SVC. They also implemented a Quantum Neural Network (QNN) architecture, which achieved 97.31% accuracy surpassing both classical CNN and the other models evaluated.

C. Research Gap

QNNs have demonstrated promising performance for image classification tasks but their application in the domain of audio signal classification is limited which paves the way for increasing performance efficiency by utilising QNNs given the intricate details captured by spectrograms.

Table I: Research Analysis

Ref. No.	CATEGORY	MAIN FINDINGS
[1][2][3]	Classical Machine Learning	The papers use and justify the use of a mel-spectrogram approach for classifying audio data using deep ANN & CNNs outperforming pre-trained ResNet & AlexNet
[4]	Classical Machine Learning	Worked on MFCC features are training data, it achieved 70.58% accuracy.
[5]	Classical Machine Learning	With spectrogram as input, ResNet with shuffle attention layers-91.27% accuracy.
[6][7][8][9]	Classical Machine Learning	The papers explored a newer approach of using gated CNNs, Audio Swin Transformers and Conformer Models
[10][11][12][13][14][15]	Quantum CNN	The papers have used a pure quantum CNN architecture achieving a better classification accuracy of 93% & 94.8% respectively than its classical counterpart.

3. Problem Definition and Objectives

This project aims to the application of a hybrid quantum convolutional neural network to the domain of audio classification utilizing different kinds of audio data including birdsong data and music genre data. Additionally, it extensively evaluates through different metrics benchmarking the performance of the proposed hybrid-QCNN model against classical CNN baselines on various audio classification tasks.

This empirical analysis aims to analyze the advantages offered by the proposed hybrid QCNN architecture in terms of computational efficiency and classification accuracy, while also shedding light on their future research directions for utilizing quantum machine learning for high-dimensional audio analysis.

4. Design of the Proposed Solution

A. Dataset Preparation:

Two datasets were used to showcase the performance of Hybrid Quantum-Classical CNNs on both small and large data sizes. The datasets utilized for this research are the GTZAN dataset a collection of 10 genres of music with 100 audio files each which is the most-used public dataset for evaluation in machine listening research for music genre recognition and the Bird song dataset collected from xeno-canto public API to explore classification/identification of a bird with an audio recording which has 1000 audio files for each bird species. For this research, only binary classification was considered for both datasets.

B. Audio Signal Analysis and Feature Extraction:

Audio signals can be represented as three-dimensional data, with time, amplitude and frequency as the three dimensions. A spectrogram is a visual representation of the frequency spectrum of a sound signal as it varies over time. For feature extraction of the audio signals, the audio signals have been converted into Mel-Spectrograms. Spectrograms are created by stacking successive Fast Fourier Transform (FFT) frames, allowing for the observation of how the magnitude or loudness of different frequency components evolves during the audio signal. An extension of the standard spectrogram is the Mel-spectrogram, which applies a nonlinear transformation to the frequency axis known as the Mel scale. The Mel scale is designed to mimic the human auditory system's perception of pitch, compressing the higher frequencies relative to the lower frequencies. By converting the frequency axis to the Mel scale, Mel-spectrograms provide a representation that more closely aligns with human auditory processing of sound. Mel-spectrograms are widely used in audio signal analysis, as visual representations to enable the extraction of salient features from complex audio data.

For this research, the audio files in both datasets were converted into Mel-Spectrograms, providing a Feature - extracted visual representation of each audio for the data.

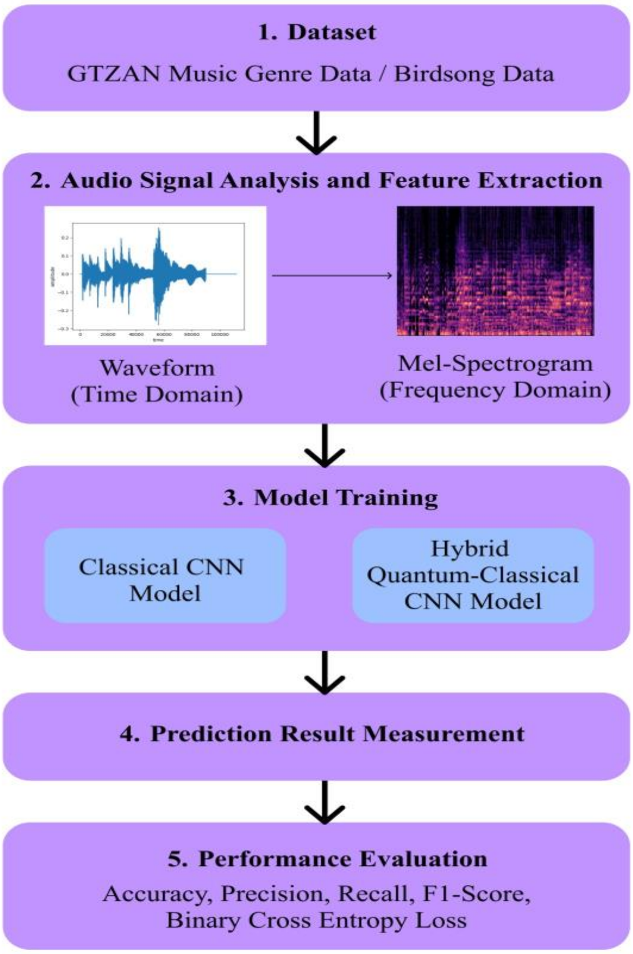

Fig. (1): System Architecture

C. Classical CNN Architecture

The CNN architecture as shown in Fig.(2) consists of three convolutional layers followed by max-pooling layers, with fully connected layers at the end for classification. At the core are convolutional layers (mathematically denoted to input data, spatial hierarchies of features, In this by Equation (1)), which systematically apply learnable filters implementation, three convolutional layers (conv1, conv2 and conv3) are employed successively, each followed by a rectified linear unit (ReLU) activation function non-linearity to the model, enabling it (to2) learn complex (mathematically denoted by Equation (1) to introduce patterns and relationships within the input spectrograms.

$$\mathrm{Y}(i,j) \;=\; (W * X)(i,j) = \sum\sum X(i+m, j+n) * W_{\ldots}(m,n) \;\ldots\; (1)$$

$$ReLU(x) = max(0, x) \dots (2)$$

Furthermore, max-pooling layers (mathematically convolutional layers, serving (3) to down sample the feature maps denoted by Equation (3)) are interspersed between the produced by the convolutional operations. This process enhances the model's translation invariance and computational efficiency by reducing the number of parameters and computations required. To mitigate over-fitting, a 2D dropout layer is incorporated after the third convolutional layer (conv3). Dropout regularization randomly zeroes out a fraction of the input channels during training, enhancing its generalization ability. Following the convolutional layers, the architecture encompasses three fully connected (dense) layers (fc1, fc2, and fc3) followed by ReLU activation functions. These layers aggregate the high-level features learned by the convolutional layers and perform classification.

$$Y(i, j) = max(X(i + m, j + n)) \dots (3)$$

The architecture is trained using stochastic gradient descent (SGD) with momentum as the optimization algorithm. SGD iteratively updates (as denoted in Equation (4)) the network parameters based on the gradients of the loss function for the parameters, gradually minimizing the discrepancy between predicted and true class distributions. Binary cross-entropy loss (mathematically denoted by Equation (7)) is employed as the optimization criterion, for its ability quantify the output based on prediction confidence.

$$w := w - \eta \nabla Q_i(w) + \alpha \Delta w \dots (4)$$

where, Q(w) is the gradient of the objective function, learning rate, η is the learning rate, α is the decay factor , w represents model's parameter vector (weights and biases).

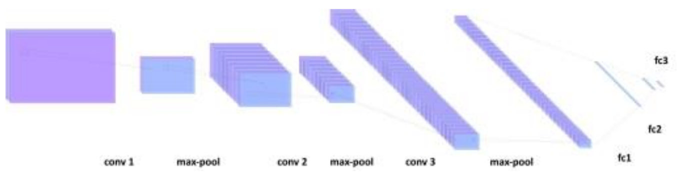

Fig. (2): Classical CNN Architecture

D. Hybrid Quantum-Classical CNN Architecture:

The classical part of the model is a convolutional neural network (CNN) comprising three convolutional layers (conv1, conv2, conv3) each followed by the ReLU activation function and

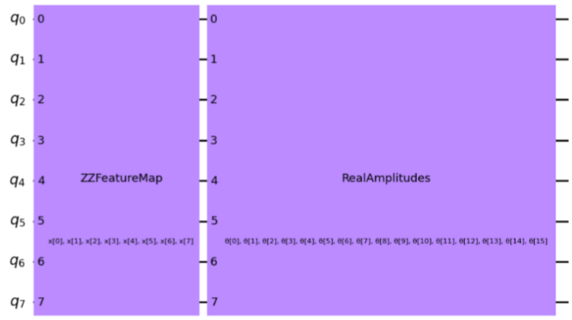

Fig. (3): Quantum Circuit 'qc'for the Quantum layer

interspersed with max-pooling layers, a dropout layer for regularization, and two fully connected layers (fc1, fc2). This CNN component is responsible for extracting high-level features from the input data through a series of convolutional operations, pooling, and non-linear activations.

The output from the second fully connected layer fc2 is passed to an EstimatorQNN layer, it takes in a parametrized quantum circuit with 8 qubits (as shown in Fig. (3)) which is composed of a FeatureMap and ansatz circuits. The parametrized quantum circuit is represented mathematically as:

$$qc = Ansatz\ (FeatureMap\ (|\ q_0\ _1\ q_2\ q_3\, q_4\, q_5\ q_6\ q_7 \rangle\ ))\ \ldots (5)$$

The EstimatorQNN takes the quantum circuit qc (denoted in Equation5) as input, the input parameters from the FeatureMap that will be treated as network inputs and weight parameters from the ansatz that will be treated as network weights. Quantum feature maps convert classical data to quantum data: $(\phi(\bar{x}))$ where $\phi(..)$ is a classical function and $V(..)$ is the parametrized circuit qc is a Second-order Pauli-Z evolution circuit. ZZFeatureMap embedding $ZZFeatureMap$ $embedding$ n -dimensional classical data on n-qubits:

$$ZZFeatureMap\ embedding = U_{\phi(x)}\ H^{\otimes n}\ \ldots (6)$$

Where, $U_{\phi(x)} = \left(i \sum_{S \subseteq [n]} \phi_s(x) \prod_{i \in S} Z_i \right)$

Z_t is a-gate on the i-th qubit.

$$\phi\,(x) = x_i, \phi_{\{i,j\}}$$

$$(x) = (\pi - x_0)(\pi - x_i)$$

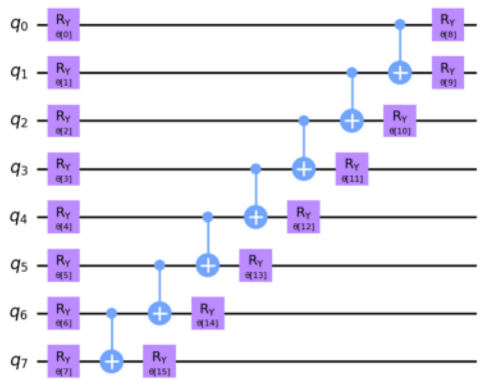

Fig. (4): Visualization of the Real Amplitudes circuit

The output from passed into an ansatz circuit. Real Amplitudes circuit is a heuristic trial wave function used as Ansatz (starting point for approximations or optimizations) in classification circuits in machine learning The circuit consists of alternating layers of Y rotations and CX entanglements. The RealAmplitudes circuit being used as an ansatz in the proposed architecture with 1 repetition on 8 qubits with full entanglement is shown in Fig. (4)
The output of the QNN layer is further processed by the final fully connected layer fc3, and the result is concatenated with its complement along the last dimension, producing a tensor with two channels that represent the predicted class probabilities.

The QNN layer is integrated into the classical CNN model using the TorchConnector class, which acts as an interface between the quantum and classical components.

5. Result Analysis

The performance of different models was evaluated using several measures, including accuracy, f1-score, precision, recall, and Cross-entropy Loss for each dataset.

$$l\,n = -\,wn\,[yn.log\;x_n + (1 - y_{\,n}).log(1 - x_n\,)\,]\;\ldots\;(7)$$

Cross-entropy penalizes greatly for being very confident and wrong. Unlike the Negative Log-Likelihood Loss, which doesn't punish based on prediction confidence, Cross-Entropy punishes incorrect but confident predictions, as well as correct but less confident predictions. Binary Cross Entropy (BCE) loss function is used since it computes the difference between two probability distributions for a provided set of occurrences (denoted in Equation (7)). It is used to work out a score that summarizes the average difference between the predicted values and the actual values. The cross-entropy loss needs to be minimized to enhance the accuracy and confidence of the model. A Model Performance Comparison for small dataset as shown in TABLE III, the Hybrid QCNN attained a training accuracy of 0.9812 comparable to its classical counterpart on the GTZAN music genre classification dataset (TABLE II), its generalization capability on the testing data was relatively poor, achieving an accuracy of only 0.875. Fig. (5). and Fig. (6). presents the epoch-wise cross entropy loss and accuracy comparison between classical and quantum-classical hybrid models in the same classification task. It shows that the quantum-classical hybrid model achieves a similar training accuracy as the classical CNN model with increasing number of iterations. Although, there is only a small decrease in the cross-entropy loss with increasing number of training iterations for Hybrid Quantum-Classical CNN model.

A. Model Performance Comparison for small dataset

As shown in TABLE III, the Hybrid QCNN attained a training accuracy of 0.9812 comparable to its classical counterpart on the GTZAN music genre classification dataset (TABLE II), its generalization capability on the testing data was relatively poor, achieving an accuracy of only 0.875.

Fig. (5). and Fig. (6). presents the epoch-wise cross entropy loss and accuracy comparison between classical and quantum-classical hybrid models in the same classification task. It shows that the quantum-classical hybrid model achieves a similar training accuracy as the classical CNN model with increasing number of iterations. Although, there is only a small decrease in the cross-entropy loss with increasing number of training iterations for Hybrid Quantum-Classical CNN model.

TABLE II: Results of GTZAN Music Genre Classification Dataset using Classical CNN Architecture

	Accuracy	Cross - Entropy Loss	F1-Score	Precision	Recall
Train	0.9875	0.0287	0.98734	1	0.975
Test	0.925	0.01374	0.9302	0.8695	1

TABLE III: Results of GTZAN Music Genre Classification using Hybrid Quantum-Classical CNN Architecture

	Accuracy	Cross - Entropy Loss	F1-Score	Precision	Recall
Train	0.9812	0.6087	0.9808	1	0.9624
Test	0.875	0.6271	0.8837	0.8261	0.9499

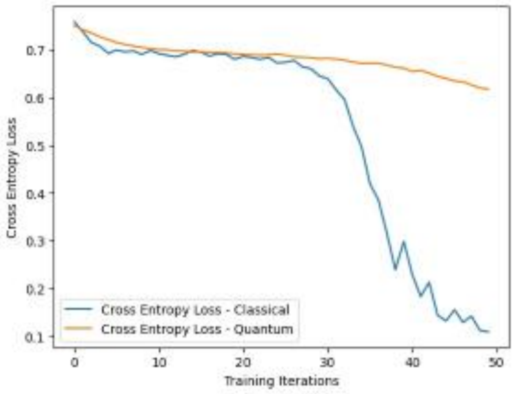

Fig. (5): Epoch-wise Cross-entropy Loss Comparison between Classical and Quantum-Classical Hybrid Models in GTZAN Music Genre Classification

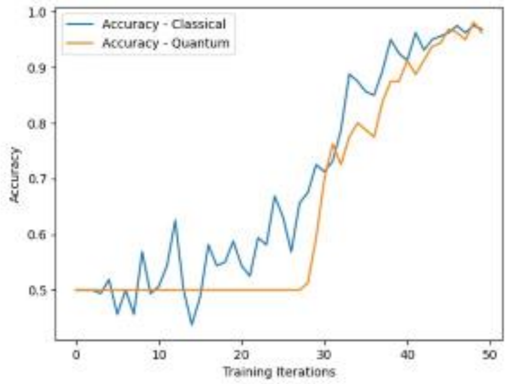

Fig. (6): Epoch-wise Accuracy Comparison in

GTZAN Music Genre Classification

TABLE V shows the improvement in the performance of the Hybrid Quantum-Classical CNN when a larger dataset was used. It achieved a training and testing accuracy of 0.9523 and 0.9504 respectively, thus surpassing its classical counterpart which had a testing accuracy of 0.9288 as seen in TABLE IV. There was also a significant decrease in cross-entropy loss with an increasing number of iterations as compared to when the smaller GTZAN dataset was used.

Fig. (7) compares the epoch-wise cross-entropy loss between classical and quantum-classical hybrid models in birdsong classification and Fig. (8) presents the epoch-wise accuracy comparison between classical and quantum-classical hybrid models in the same classification task.

TABLE IV: Results of Birdsong Classification Data using Classical

CNN Architecture

	Accuracy	Cross - Entropy Loss	F1-Score	Precision	Recall
Train	0.9561	0.1966	0.9585	0.9763	0.9413
Test	0.9288	0.0005486	0.9317	0.9657	0.8999

TABLE V: Results of Birdsong Classification Data using Hybrid Quantum-Classical CNN Architecture

	Accuracy	Cross - Entropy Loss	F1-Score	Precision	Recall
Train	0.9523	0.5529	0.9559	0.9535	0.9583
Test	0.9504	0.4276	0.9537	0.9595	0.948

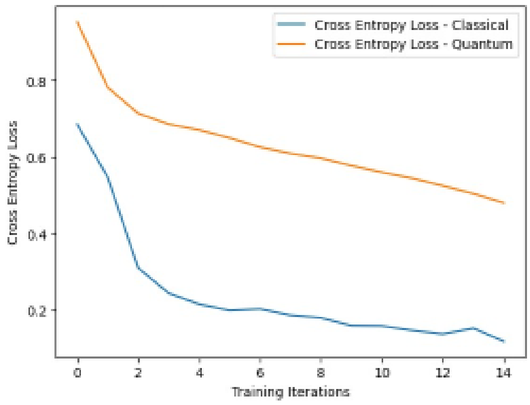

Fig. (7): Epoch-wise Cross-entropy Loss Comparison in Birdsong Classification

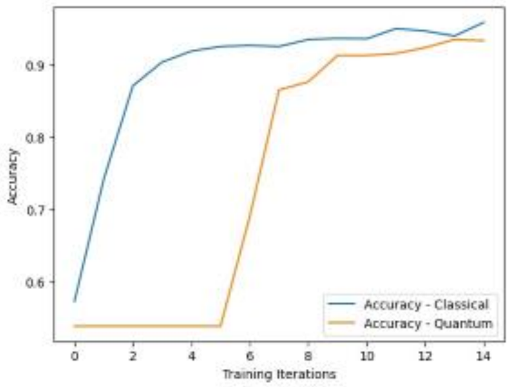

Fig. (8): Epoch-wise Accuracy Comparison in Birdsong Classification

C. Result Analysis

The results demonstrated that the hybrid quantum-classical architecture exhibited a reduction in overfitting on the larger dataset compared to the classical CNN, consequently yielding a significant increase in test accuracy.

A notable observation was the positive correlation between overall accuracy and dataset size for the quantum-classical hybrid model as shown in Fig.(9). As the volume of training data increased, the hybrid architecture exhibited a corresponding rise in overall classification accuracy. Concurrently, the hybrid approach demonstrated a substantial reduction in over-fitting when compared to its performance on the smaller dataset.

An elevated cross-entropy loss implies a higher degree of uncertainty or inaccuracy in the predicted probabilities for individual instances. The high cross-entropy loss and good accuracy suggest that the proposed hybrid model is performing well in terms of overall classification accuracy, but is struggling with confidence or precision in its probability estimates for each specific data point.

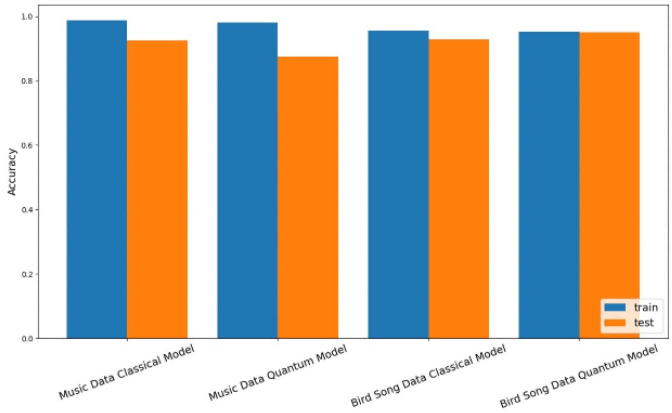

Fig. (9): Comparison of Test and Train Accuracy for GTZAN Music Genre Classification and Birdsong Classification Data

6. Conclusion and Scope of Future Scope

This study investigated the application of hybrid quantum classical convolutional neural networks (QCNNs) for audio classification tasks utilizing Mel-spectrogram representations. The proposed hybrid QCNN architecture, integrating a classical CNN for feature extraction with a quantum neural network layer, was evaluated on two audio datasets of varying sizes. For the smaller GTZAN dataset, the hybrid QCNN model achieved comparable training accuracy (0.9812) to the classical CNN baseline (0.9875). However, it struggled to generalize effectively to the test data, attaining a lower accuracy of 0.875compared to 0.925 for the classical CNN. The hybrid model also exhibited significantly higher cross-entropy loss, indicating difficulties in calibrating prediction confidence. When evaluated on the larger birdsong dataset, the hybrid QCNN demonstrated improved performance surpassing the classical CNN. It achieved higher test accuracy of 0.9504 compared to 0.9288 for the classical model, while also mitigating over-fitting issues. This empirical evidence highlights the potential advantages offered by quantum neural network architectures in handling high-dimensional and complex audio data. Overall, this study contributes to the nascent field of quantum machine learning for audio analysis by demonstrating the application of hybrid quantum-classical architectures. The findings underscore the potential computational benefits of leveraging quantum effects for complex audio classification tasks. There is a lack of interpretability and explainability in QNNs due to the intricate nature. Future work could focus on developing quantum-specific explainability techniques to provide insights into the decision-making processes of QCNNs.

Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)

Publications

Title: 15th ICCCNT 2024 (15th International IEEE Conference on Computing Communication and NetworkingTechnologies)

Organised By : IIT-Mandi

Venue : IIT-Mandi

Date : 18-28 June,2024

Status : Submitted

15th ICCCNT 2024 submission 2323 Inbox ×

15th ICCCNT 2024 <15thicccnt2024@easychair.org>
to me

Dear authors,

We received your submission to 15th ICCCNT 2024 (15th International IEEE Conference on Computing Communication and Networking Technologies):

Authors : Yash Thakar, Bhuvi Ghosh, Vishma Adeshra and Dr. Kriti Srivastava
Title : Performance Analysis of Hybrid Quantum-Classical Convolutional Neural Networks for Audio Classification
Number : 2323
Track : Deep Learning

The submission was uploaded by Yash Thakar <yashthakar2710@gmail.com>. You can access it via the 15th ICCCNT 2024 EasyChair Web page

https://easychair.org/conferences/?conf=15thicccnt2024

Thank you for submitting to 15th ICCCNT 2024.

Best regards,
EasyChair for 15th ICCCNT 2024.

References

[1] Z. Gao, T. Liu, M. Zhu, J. Li, Y. Ning and Z. Wang, "Environmental Sound Classification Using CNN Based on Mel-spectogram," 2023 2nd International Conference on Artificial Intelligence and Blockchain Technology (AIBT), Zibo, China, 2023, pp. 41-45, doi: 10.1109/AIBT57480.2023.00015. keywords: {Deep learning;Training data;Feature extraction;Data processing;Blockchains;Convolutional neural networks;Data mining;CNN;Environmental sound classification;Melspectogram},

[2] R. Pillai, N. Sharma, D. Upadhyay, S. Dangi and R. Gupta, "Sonic Signatures: Sequential Model-driven Music Genre Classification with Mel Spectograms," 2024 Fourth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 2024, pp. 1-6, doi: 10.1109/ICAECT60202.2024.10468856. keywords: {Training;Measurement;Deep learning;Autonomous systems;Computational modeling;Data models;Spectrogram;Deep Learning;Sequential Model;Music Genre;Education;Innovation;Practical Training},

[3] L. Le, A. N. M. H. Kabir, C. Ji, S. Basodi and Y. Pan, "Using Transfer Learning, SVM, and Ensemble Classification to Classify Baby Cries Based on Their Spectrogram Images," 2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems Workshops (MASSW), Monterey, CA, USA, 2019, pp. 106-110, doi: 10.1109/MASSW.2019.00028. keywords: {Pediatrics;Support vector machines;Spectrogram;Machine learning;Predictive models;Databases;Convolutional neural networks;Baby cry classification, spectograms, Resnet, SVM, decision fusion},

[4] V. Bansal, G. Pahwa and N. Kannan, "Cough Classification for COVID-19 based on audio mfcc features using Convolutional Neural Networks," 2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, India, 2020, pp. 604-608, doi: 10.1109/GUCON48875.2020.9231094. keywords: {COVID-19;Sensitivity;Labeling;Convolutional neural networks;Mel frequency cepstral coefficient;Viruses (medical);Spectrogram;Cough Classifier;COVID-19;MFCC (mel frequency cepstral coefficients);CNN (convolutional neural networks);Deep Learning},

[5] D. Kinha, G. Gulati, A. Baranwal and D. K. Vishwakarma, "Classifying Heart Sounds for Disease Detection Using Deep Learning Methods," 2023 4th International Conference for

Emerging Technology (INCET), Belgaum, India, 2023, pp. 1-6, doi: 10.1109/INCET57972.2023.10170444. keywords: {Heart;Deep learning;Analytical models;Perturbation methods;Neural networks;Data models;Recording;Heart Sound Classification;Feature Selection;PhysioNet;PASCAL;ResNet;ResNeXt;Spectogram;Shuffle Attention},

[6] Y. Xu, Q. Kong, W. Wang, and M. D. Plumbley, "Large-Scale Weakly Supervised Audio Classification Using Gated Convolutional Neural Network," in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2018. doi: 10.1109/ICASSP.2018.8461975.

[7] Y. Li, X. Li, Y. Zhang, W. Wang, M. Liu, and X. Feng, "Acoustic Scene Classification Using Deep Audio Feature and BLSTM Network," in ICALIP 2018 - 6th International Conference on Audio, Language and Image Processing, 2018. doi: 10.1109/ICALIP.2018.8455765.

[8] Zhao, Hang & Zhang, Chen & Zhu, Belei & Ma, Zejun & Zhang, Kejun. (2022). S3T: Self-Supervised Pre-training with Swin Transformer for Music Classification.

[9] Anmol Gulati and James Qin and Chung-Cheng Chiu and Niki Parmar and Yu Zhang and Jiahui Yu and Wei Han and Shibo Wang and Zhengdong Zhang and Yonghui Wu and Ruoming Pang,Conformer: Convolution-augmented Transformer for Speech Recognition,2020,2005.08100,arXiv,eess.AS

[10] G. Chen, Q. Chen, S. Long and W. Zhu, "Quantum Convolutional Neural Network For Image Classification," 2020 8th International Conference on Digital Home (ICDH), Dalian, China, 2020, pp. 116-120, doi: 10.1109/ICDH51081.2020.00028. keywords: {Image recognition;Quantum entanglement;Convolution;Qubit;Feature extraction;Breast cancer;Convolutional neural networks;Quantum Convolutional Neural Network;MERA Circuit;Image Classification;TensorFlow Quantum},

[11] V. Rajesh, U. P. Naik and Mohana, "Quantum Convolutional Neural Networks (QCNN) Using Deep Learning for Computer Vision Applications," 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), Bangalore, India, 2021, pp. 728-734, doi: 10.1109/RTEICT52294.2021.9574030. keywords: {Deep learning;Computers;Training;Computer vision;Quantum computing;Image recognition;Neural networks;Quantum Convolutional Neural Network (QCNN);Computer Vision;FPGA;Video Surveillance;Deep Learning;Machine Learning},

[12] F. Fan, Y. Shi, T. Guggemos and X. X. Zhu, "Hybrid Quantum-Classical Convolutional Neural Network Model for Image Classification," in IEEE Transactions on Neural Networks and Learning Systems, doi: 10.1109/TNNLS.2023.3312170.keywords: {Quantum computing;Computational modeling;Quantum state;Qubit;Integrated circuit modeling;Logic gates;Image classification;Image classification;quantum circuit;quantum machine learning (QML);remote sensing imagery},

[13] X. Chen, B. Jiang, Y. Wu and Y. Ye, "A QCNN Algorithm Deployed on A Quantum Computer," 2022 International Conference on Computing, Communication, Perception and Quantum Technology (CCPQT), Xiamen, China, 2022, pp. 390-395, doi: 10.1109/CCPQT56151.2022.00074. keywords: {Analytical models;Quantum computing;Image recognition;Computational modeling;Optimization methods;Quantum state;Data models;Quantum Computing;Quantum Convolutional Neural Network;Image Recognition;Machine Learning},

[14] M. N. Sachane and S. A. Patil, "Adaptive Spotted Hyena Optimizer-enabled Deep QNN for Laryngeal Cancer Classification," 2022 International Conference on Edge Computing and Applications (ICECAA), Tamilnadu, India, 2022, pp. 1025-1032, doi: 10.1109/ICECAA55415.2022.9936500. keywords: {Adaptation models;Head;Filtering;Neural networks;Surgery;Sensitivity and specificity;Feature extraction;Laryngeal Cancer (LCA);Laryngoscopy;Spider Local image feature (SLIF);Spotted Hyena Optimizer (SHO);Deep Quantum Neural Network (Deep QNN)},

[15] S. Prabhu, S. Gupta, G. M. Prabhu, A. V. Dhanuka and K. V. Bhat, "QuCardio: Application of Quantum Machine Learning for Detection of Cardiovascular Diseases," in IEEE Access, vol. 11, pp. 136122-136135, 2023, doi: 10.1109/ACCESS.2023.3338145. keywords: {Computational modeling;Quantum computing;Cardiovascular diseases;Feature extraction;Biomedical imaging;Static VAr compensators;Electrocardiography;Quantum machine learning (QML);quantum support vector machine (QSVM);Pegasos;quanvolutional neural network (QNN);medical image classification;cardiovascular disease classification},

PAPER NAME

Quantum_IPD_Project_report.pdf

WORD COUNT	CHARACTER COUNT
5318 Words	33719 Characters
PAGE COUNT	FILE SIZE
29 Pages	1.0MB
SUBMISSION DATE	REPORT DATE
May 5, 2024 6:37 PM GMT+5:30	May 5, 2024 6:38 PM GMT+5:30

9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 5% Internet database
- 4% Publications database
- Crossref database
- Crossref Posted Content database
- 5% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less then 8 words)
- Manually excluded text blocks