Design and development of an 'early prediction machine' for Colorectal Cancer from pathological images through quantum image processing technique – a theranostic approach

Rohith V.[1], P. K Krishnan Namboori[2],
{rohithvpkd@gmail.com[1], n_krishnan@cb.amrita.edu}

Department of Computer Science, Amrita School of Arts & Sciences, Mysore Campus[1],
AMMAS Research Lab, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeethom[2].

Abstract. The Cancer has been reported as a major terminal disease of the world and the number of deaths due to various types of cancer is increasing day by day. More than 30% of the death due to cancer is due to colorectal cancer (CRC) resulted by mutations in the WNT signalling pathway. In most of the cases, early detection of the disease and proper treatment may help in resulting complete cure. Improvements in modern technology with deep neural networks and artificial learning along with image processing enables diagnosis and detection of cancer cells in the early stages. In the present work, the possibility of using _quantum processing technique or 'Qubit computing' has been explored to classify malignant and benign cells. The dataset used is pathological images of colorectal cancer processed using a 2-bit quantum circuit. The processing has been carried out using 'IBM Quantum computing (IBM-Q)'. Even with a small dataset and with 4-qubit platform, more than 50% accuracy has been observed. Higher percentage of accuracy may be obtained by optimizing the number of qubits and by using bigdata.

Keywords: Colorectal cancer, Hybrid quantum model, Artificial Intelligence, Deep learning.

1 Introduction

The colorectal cancer (CRC) has been identified as a common type of cancer seen among most of the cancer patients and is reported as the third terminal disease. In most cases the disease is detected only in the well-advanced final stages, making a complete cure hardly possible.Moreover, most of the patients would not have any major external symptoms in the beginning. The modification in the WNT signalling pathway has been reported as the major cause of the disease resulted by mutations in the Adenomatous polyposis coli (APC) gene. The most popular method for early detection of CRC is Computer Aided Diagnosis (CAD) following colonoscopy technique. The CAD assistance covers the entire image processing stages such as segmentation, screening, feature extraction, colour coding of polyps, visualization [1] and decision making to predict CRC. In the early detection of colorectal cancer, the geometric shape of the polyp is considered as the major attribute for the decision making. In the 'flow cyclometry' technique, machine learning [2] methods have been used. In most of these cases, the classification of polyps was done by physical examination of data to

train the classifier and then create a prediction model to classify the cells as malignant or benign [3].

The development of deep learning technique changed the entire scenario, opening deep Convolution Neural Networking as a potential machine learning technique for addressing biological problems. Recently, quantum computing platform has been specified as the most suitable computing environment for bigdata analysis, especially biological datasets. However, the exact number of qubits to be used in the classification problem depends on the nature of dataset to be used. Different quantum computing environments have been suggested.

Maria Schuld et. al. [4] describes an 'input output relation' from examples to illustrate new objects in machine learning with quantum computing. Distributed secure quantum machine learning by Yu-Bo Scheng et. al. [5] describes the remote quantum machine learning with a classical client with quantum technology, where data privacy is preserved. Quantum algorithm for supervised and unsupervised machine learning by Seth Lloyd et. al. [6] describes about the learning methods and algorithms for cluster assignment and cluster finding. Misha et. al. [7] describes about analysis on early detection tool using pharmacogenomic as well as one shot learning algorithm of deep learning technique. Vaisali et. al. [8] proposes a detection machine that addresses the problem faced by expensive conventional mutation detection technique and make them more individual specific. Hima et. al. [9] describes the common pharmacogenomic aspects associated with CRC, especially for the Indian population. Peter Wittek [10] explains the abstract trend in quantum computing and the applied research on machine learningcovering the classical learning methods in quantum computing and its complexity. A quantum circuit base for quantum computation uses a sequence of quantum gates, which are reversible[13][4].

The reversible gate on analogous structure is termed as n-qubit register. A reversible n-qubit quantum gate has unitary mapping onto itself. The computational methods in quantum circuits are characterized by linear algebra over Hilbert space, where the number of input lines matches the number of output lines. A hybrid quantum model with a variational working classifier is used in this work as depicted in Fig 1.
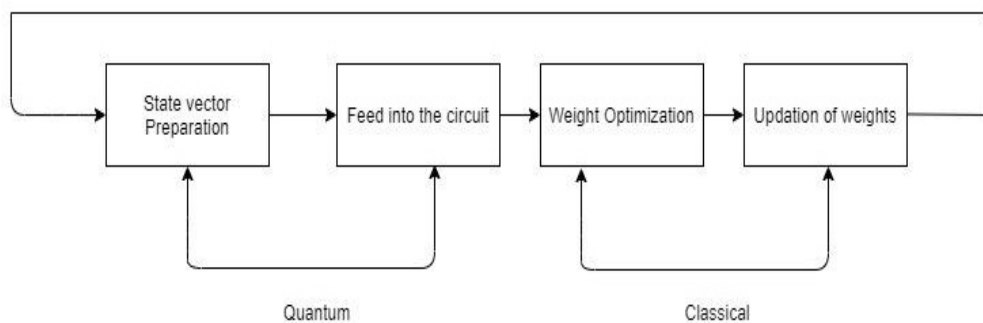
Fig. 1. Hybrid Quantum Model.

2 Hybrid Quantum Model

The variational classifier model used in the quantum circuit is a feed forward neural network, consisting of reversible parameters. The circuit consists of multiple layers of network and each individual layer in the circuit will undergo a set of operations. All operations carried out in the individual layers of network are performed by quantum gates. Variational classifier is visualised in Fig. 2.
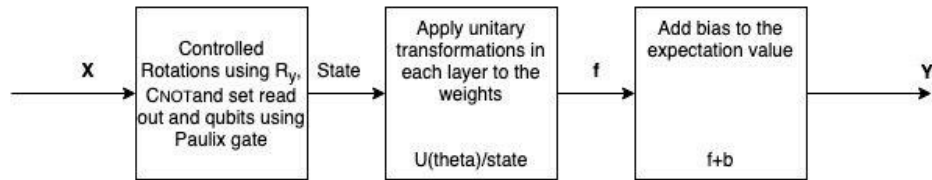

Fig. 2. Variational Classifier.

The variational circuit implemented in this work is collected from Mottonen et. al. [11] and Nielsen et. al. [12]. The cross-platform Python library, PennyLane has been used for the quantum machine learning, automatic differentiation, and optimization of hybrid quantum-classical computations of the work.

3 Implementation

The CRC microscopic images were collected from ISIC archive colorectal cancer dataset. These images comprise of 3 channel (RGB) with a size of 128x128 pixels. Conventional deep learning technique called autoencoder, that reduces the dimensionality, is used due to high resolution and dimensionality of the samples. The procedure is shown in Fig 3 and a sample image is included in Fig.4.
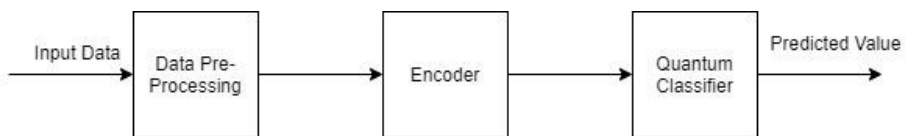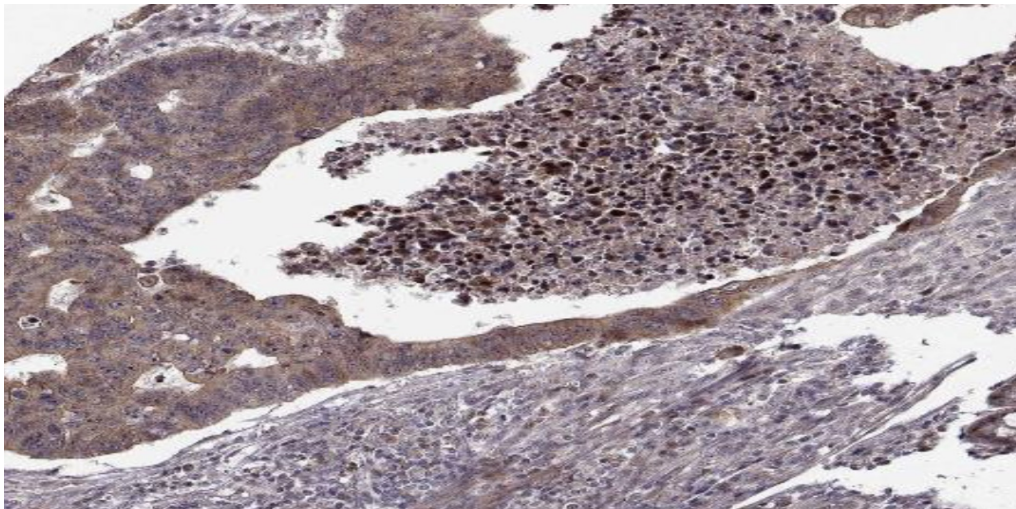

Fig. 3. Procedure.

Fig. 4. Dataset Image.

For training purpose 80% images per class is taken and the remaining 20% for the validation. In data pre-processing phase, each of the biological image is compressed to a size of 64x64 as they are less prone to noise while feeding into autoencoder circuit. Since these images are of very high resolution, they need to be encoded to lower resolution before fitting into the quantum circuit. The autoencoder is a variational classifier model in the neural network architecture to reconstruct the input that makes it fit for the representation of data. Autoencoder consists of an encoder and decoder phase as depicted in Fig 5. The variational autoencoder is implemented in such a way that it reconstructs the data with minimal errors. This can be achieved by characterizing the data in a probabilistic approach. Pre-processing of the data can be done by using the encoder section, where it is called as recognition model and the decoder section referred as generative model.
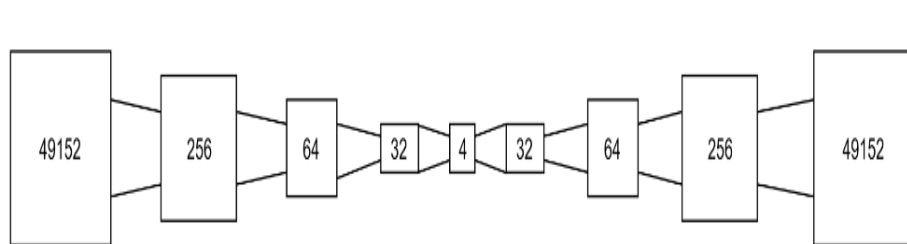

Fig. 5. Encoder - Decoder.

4 AutoEncoder

The encoder phase pre-process the data by compressing the input data or downscaling it to a certain number of dimensions. Decoder phase reconstruct the encoded image into the original dimensions. During the training stage, the encoder section extracts vital information from the input data which is fed into the quantum circuit and the decoder section reconstruct these encoded images into realistic ones. The proposed hybrid quantum model uses a popular deep learning framework called PyTorch to encode the images. These images are downscaled into 64x64 vector of 2 components in encoder section. These vectors are realized into quantum amplitudes and fed into the hybrid quantum circuit as input data.

5 Results and Discussions

The proposed hybrid quantum circuit used in this approach is trained to perform binary classification task and is carried out with an average accuracy more than 50%, which is a very appreciable level of accuracy for a biological database with a very minimum number of samples. Only, 242 images of APC and their associated genes obtained as a result of microscopic analysis after biopsy were used as input images in training, testing and validation process. The input data was encoded and trained for 60 epochs with a batch size of 32 and the weights and biases are sampled randomly to form a uniform distribution. Since the input data is encoded to a very low dimensional space, the accuracy of the data is attributed due to the loss in the data. The major improvement in accuracy can be attained by optimizing the number of qubits. Accuracy may also be improved by manual feature selection methods during pre-processing that could better train the machine. The main goal of the approach was to analyse a very small quantum classical hybrid model, which can be used for ease of interference at a very low budget.

6 Conclusion

The cutting-edge method called 'quantum computing' was applied for the early detection of colorectal cancer. Various studies have been reported using deep learning, bigdata and neural networks in this regard. The present work aims at seeking the possibility of using 'Quantum qubit computing technique' for biological datasets. The Computing environment for a bigdata set has been fully designed and developed. On trial run using a small dataset, more than 50% accuracy has been identified. The improvement in number of qubits and manual selection of dimensions during pre-processing of images could gradually improve the accuracy. The same strategy can also be used for a multi-class problem.

References

[1] Sareena, Ajay Mittal, ManjeetKaur, 'Computer-aided-Diagnosis in Colorectal Cancer: A Survey of state of the art techniques', IEEE Explore, 2017 Euromicro Conference on Digital System Design (DSD), 30 September 2017.

[2] S. Takeda et. al, ―Flow cytometry as a diagnostic method for colorectal cancer', IEEE Explore, 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, November 2012.

[3] Chong Li et. al, ―Improving colorectal polyp classification based on physical examination data ―an ensemble learning approach', IEEE Explore, 2017 13th IEEE Conference on Automation Science and Engineering (CASE), January 2018.

[4] Schuld, Maria, IlyaSinayskiy, and Francesco Petruccione. ―An Introduction to Quantum Machine Learning.' Contemporary Physics 56.2 (2014): 172–185. Crossref. Web.

[5] Yu-Bo Sheng, Lan Zhou, Distributed secure quantum machine learning, In Journal of Science Bulletin, Volume 62, Issue 14, 2017, Pages 1025-1029, ISSN 2095-9273.

[6] Seth Lloyd, MasoudMohseni, Patrick Rebentrost, Quantum algorithms for supervised and unsupervised machine learning, arxiv preprint:1307.0411, 2013.

[7] Misha Abraham, A.M. HimaVyshnavi, ChungathSrinivasan and Dr.P.K. Krishnan Namboori (2019) Healthcare security using blockchain for pharmacogenomics, Journal of International Pharmeceutical Research (46)-1: 529-533.

[8] B.Vaisali, P.K. Krishnan Namboori.(2019)Tumor Hypoxia Mutation Detection Using Deep Learning-A Deep Drug Designing Strategy, International Journal of Scientific & Technology Research Volume 8, Issue 11.

[9] Am, Hima&Namboori, Dr. Krishnan & O M, Deepak & . C, Lakshmi. (2017). Evaluation of Colorectal Cancer (CRC) Epidemiology A Pharmacogenomic Approach. Journal of Young Pharmacists. 9. 36-39.

[10] Wittek, Peter. (2014). Quantum Machine Learning: What Quantum Computing Means to Data Mining.

[11] Mikko M ottonen,∗Juha J. Vartiainen, Ville Bergholm, and Martti M. Salomaa, Transformation of quantum states using uniformly controlled rotations, 1 Feb 2008.

[12] Michael A. Nielsen, Isaac L. Chuang, Quantum Computation and Quantum Information: 10th Anniversary Edition, 2011 Book. Variational Autoencoders. Towards Data Science, 2018. [Online]. Available: https://www.jeremyjordan.me/variational-autoencoders [Accessed: 08-November- 2019].

[13] Vinod Jagannath Kadam, Shivaji rao Manik rao Jadhav, K.Vijayakumar, "Breast Cancer Diagnosis Using Feature Ensemble Learning Based on Stacked Sparse Auto encoders and Soft max Regression", Image & Signal Processing, springer, june 2019.

[14] K. Vijayakumar , K. Pradeep Mohan Kumar ,Daniel Jesline, "Implementation of Software Agents and Advanced AoA for Disease Data Analysis", journal of medical systems, Part of Springer Nature 2019.