GMR
The Original

Hybrid Quantum–Machine Learning Models For High-Dimensional Gene Expression Profiling And Early Risk Prediction Of Parkinson's And Alzheimer's Disease

Dr. S Beula Princy[1] , Mary Jacob[2] , Sarala G[3] , Dr. S.Rajanaryanan[4] , Dr. S.Kannan[5] , R. Naveenkumar[6], Deepender[7]

[1]Assistant Professor, Department of Information Technology, PSGR krishnammal College for Women, Coimbatore, Tamilnadu, India. Email: sharmi1318@gmail.com https://orcid.org/0009-0008-4138-027X
[2]Dept. of Computer Science, Kristu Jayanti (Deemed to be University), Bengaluru, India. Email: maryjacob05@gmail.com ORCID: 0000-0003-4016-3544
[3]Professor, Department of Neurology, Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research,Chennai, Tamilnadu, India. Email: saralag@maher.ac.in
[4]Professor, Department of Computer Science and Engineering, Vinayaka Mission's Kirupananda Variyar Engineering College, Salem(Vinayaka Mission's Research Foundation), Tamilnadu, India. Email: rajanarayanan@vmkvec.edu.in 0009-0007-4219-4084
[5]Assistant Professor, Department of Biomedical Engineering, Vinayaka Mission's Kirupananda Variyar Engineering College, Salem (Vinayaka Mission's Research Foundation), Tamilnadu, India. Email: kannan@vmkvec.edu.in 0000-0002-4648-9804
[6]Dept of CSE, School of Engineering and Technology, CGC University Mohali-140307, Punjab India. Email: drnk1983@gmail.com 0000-0001-9033-9400
[7]Assistant Professor, Faculty of Computing, Guru Kashi University, Bathinda, Punjab, Email:- deependerduhan6@gmail.com Orcid: 0000-0002-0529-4007

Abstract

Molecular changes in the early stage of Parkinson's disease (PD) and Alzheimer disease (AD) occur before disease onset, and thus, transcriptomic profiling could be an effective approach to risk prediction in the early phases of the disease. Nonetheless, the generated datasets of gene expression are very high-dimensional usually having thousands of genes with few samples, that are not readily accommodated by traditional machine learning methods, which face the risk of overfitting and low nonlinear representational properties. Strong computational models are thus needed to obtain discriminative molecular signatures and generalise at the same time. This work suggests a quantum-machine learning (QML) framework to profile high-dimensional gene expression to enhance prediction of the early risk of both PD and AD and maintain biological insights. Most public transcriptomic datasets were preprocessed with log transformation, Z-score normalisation, and then they were subjected to the analysis of differential expression and mutually informative feature selection. The encoding of the genes of interest into a variational quantum circuit was through angle encoding, which allowed nonlinear representation to a higher-dimensional feature space. Regularised cross-entropy loss was used to train a hybrid quantum -classical architecture that used parameterized quantum layers followed by a classical classifier. Accuracy, precision, recall, F1-score and ROC-AUC were used to measure performance that was benchmarked against support vector machines, random forests and classical neural networks. This model proposed was more successful in higher ROC-

AUC, F1-scores in single datasets of PD, and AD, and showed greater ability to perform in high-dimensional and small samples. The process of neuroinflammation via genes associated with neuroinflammatory pathways, mitochondrial dysfunction, and synaptic signaling were determined by feature importance analyses and pathway enrichment as per known neurodegenerative mechanisms.

Keywords Hybrid quantum–machine learning, Gene expression profiling, Parkinson's disease, Alzheimer's disease, Neurodegenerative risk prediction, Variational quantum circuits.

1. INTRODUCTION

Parkinson and Alzheimer are two of the most common debilitating forms of neurodegenerative diseases that cause significant clinical, social and economic challenges globally. The two maladies are distinguished by a gradual neuronal degeneration which eventually gives rise to motor impairment in PD and cognitive impairment in AD. The major limitation in the present day clinical practise is that the decisive diagnosis usually happens when the neurons have already suffered a substantial amount of harm. Mounting evidence suggests that molecular and transcriptomic changes begin to appear several years (and even decades) prior to the occurrence of apparent clinical symptoms [4,18]. Such initial changes in the molecules involve the maladaptation of genes implicated in neuroinflammation and mitochondrial disturbance, oxidative stress, synaptic signalling, and aggregation of proteins [7,14]. Therefore, the process of discovering dependable transcriptomic biomarkers, which could reveal early pathological indicators, has become the priority of the fields of translational neuroscience and precision medicine. Gene expression profiling would provide an exceptional arena of capturing these delicate, but biologically meaningful alterations so as to risk stratify and possibly engage early intervention measures [6,12]. Although promising, data on modelling high-dimensional gene expression offers significant challenges in analysis. Transcriptomic measurements normally include thousands to tens of thousands of genes on a relatively small sample population, resulting in what is commonly termed as the curse of dimensionality. The number of features may be very large in such scenarios compared to the number of observations, and thus the number of observations is more likely to overfit the model and decrease the model generalizability [11]. Small xenoselect samples limit the statistical strength and make it difficult to discover consistent molecular patterns. What is more, gene regulatory processes are characterized by complex nonlinear interactions, epistatic interactions and pathway dependencies, which cannot be readily represented by simple linear or shallow machine learning models [13]. Even though classical algorithms, including support vector machines, random forests, and deep neural networks, have proven useful in the process of genomic classification [10], they are prone to degradation in the face of high dimensionality extremes and scarcity of data. The hybrid quantum -machine learning (QML) methods provide an alternative possibility by exploiting the representational power of quantum feature spaces to generate better nonlinear representation [3]. Using variational quantum circuit theory, it is possible to embed classical data in high-dimensional Hilbert spaces, which can make the boundaries of the decisions between disease and control classes richer [2]. The combination of quantum layers and classical optimization and output units in hybrid architectures is able to merge the strengths of the two paradigms and be compatible with existing computational infrastructure [1,5]. Nonetheless, the scope of utilisation of hybrid QML frameworks in early risk prediction of neurodegenerative diseases using transcriptomics is poorly investigated. This paper will fill this gap by presenting a new hybrid quantum-classical architecture specifically for high-dimensional gene expression-analysis. The model also uses an optimised feature selection strategy, which is based on quantum encoding, quantum benchmarking of key classical models, and an overall biological interpretability cheque to make sure that predictive performance increases are within the domain of known molecular pathways.

2. Background and Related Work

Genome-wide or gene expression profiling is emerging to be an essential methodology used to study molecular changes in neurodegenerative disorders like Parkinson disease (PD) and Alzheimer disease (AD). It is indicated that in transcriptomic studies, there is extensive dysreorganization of genes with roles in neuroinflammation, and mitochondrial bioenergetics, oxidative stress response, synaptic transmission, protein aggregation, and apoptosis [4,7]. Gene expression associated with viability and survival of dopaminergic neurons and malfunctioning of the mitochondrial complex I has also been broadly recorded in PD [7]. Transcriptomic signatures to be used in AD often have an accent on amyloid precursor protein processing, tau pathology, immune activation, and impaired synaptic plasticity [14,18]. Microarray analysis and RNA sequencing, can be considered high-throughput technologies that have facilitated the complete profiling of these changes and produced massive datasets that can offer insight into disease aetiology [12]. To handle these complexities, a large set of machine learning approaches have been used on large-dimensional genomic data. Classical algorithms such as support vectors machine, random forests as well as k nearest neighbour and deep neural networks have verified different levels of achievement in the disease classifications and biomarkers identification [11,13]. Variance filtering and regularisation-related algorithms are some of the feature selection approaches that have been used to select the features and enhance the stability of the models. With these improvements, traditional machine learning frameworks are usually constrained due to overfitting especially in the case of a large number of genes relative to the sample size. Deep neural networks are strong, but they need massive data to be trained reliably and might be incapable of effective generalisation in small-sample transcriptomic research [10].

Quantum machine learning (QML) represents a new model of computations that aims to use the laws of quantum mechanics to represent data and optimise them more effectively [3]. Parameters The approach of variational quantum circuits (VQC) is based on the use of parameterized quantum gates and entanglement operations to project classical input data into high-dimensional Hilbert spaces, and may generalize nonlinear transformations [2,5]. Alternative methods known as quantum kernel methods compute the similarities in quantum feature spaces, which is theoretically beneficial in intricate classification problems [9,16]. Hybrid quantum classical learning Hybrid quantumtypes of quantum circuits use an interface with classical preprocessing and optimization layers, so they can be integrated with existing machine learning processes, and take advantage of quantum feature mappings [1,18]. The expressive power and the behaviour of generalisation of quantum neural networks has been investigated more and more in theory [1]. However, the application of QML to neurodegenerative transcriptomic analysis is sparsely developed despite the increase in interest in QML. Little effort has been made on hybrid quantum-machine learning models to predict early risk in the case of PD and AD using gene expression. Additionally, the current methods of computation usually put into focus predictive performance but neglectfully offer biological interpretability that is absolutely crucial in promoting translational relevance [17]. This breach impels the need to introduce integrative frameworks, capable of improving the accuracy of classification as well as maintain significant relationships to their associated molecular processes.

3. Materials and Methods

3.1 Dataset and Preprocessing

Transcriptomic experiments that are publicly available were utilized to create and test the proposed hybrid quantum-machine learning framework. The library of expression profiles was taken out of known repositories (like Gene Expression Omnibus (GEO)) that have standardised datasets of microarray and RNA-seq on a standard set of experimental conditions. The chosen datasets consisted of cohorts of Parkinson's disease (PD) and Alzheimer's disease (AD) and a control group of healthy individuals so that they could provide clinically appropriate disease-control differentiation. The total sample size, distribution of classes and gene dimensionality were noted down in every dataset. Transcriptomic datasets are readily high-dimensional, that is, it has many genes and a small sample of samples, a small-n, large-p situation,

and is therefore more prone to overfitting and a wobbly model estimator. Older gene expression matrices tend to be skewed because of biological variation and technical effects including probe intensity differences or sequences depth. A log transformation was performed before normalisation to stabilise the variance and reject the importance of few extreme values. This change enhances the numerical stability, as well as increases with the compatibility with the downstream optimization processes. After transformation, normalisation of gene-wise Z-score was done across samples to bring the features at similar scales. This is a necessary step since both the classical classification algorithms and the quantum coding algorithms are input sensitive algorithms. Standardisation will make sure that no single gene has a disproportionate effect on the learning process and convergence behaviour during the model training. The normalisation process of the Z-score is described in Equation (1):

$$X'_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j}, \quad \textbf{(1)}$$

X_{ij} in Equation (1) is the original value of expression of gene j in sample i, mu j is the average expression of gene j calculated in all samples, and σ_j is the standard deviation. The normalised value X_{ij} is hence the standardised variation of every expression value over its own gene-specific mean. Equation (1) is used with the objective of eliminating the scale effect that is brought about by the genes to convert all transcriptomic features into a common scale with a mean of about zero and variance of one. In addition to improving model resilience, minimizing optimization instability, and guaranteeing dependable downstream feature choice and quantum encoding it improves the strength of models. This step is fundamental in the setting of high-dimensional neuro degradative transcriptomics relative to stable hybrid quantum classical model training and generalizable risk prediction.

3.2 Feature Selection and Quantum Encoding

Since transcriptomic data is extremely dimensional, the noise reduction, discriminative power increase, and compatibility with the constraint of quantum encoding used a multi-stage approach on the feature selection strategy. Quantum computing Thousands of genes cannot be directly encoded into quantum circuits on current quantum hardware or simulators because there is a lack of qubits and depth constraints in circumvention. Thus dimensionality reduction presented was done in an organised and biologically informed way. To begin with, differentiation expression was done to determine the genes that had statistically significant difference in their expression between the disease and control groups. Depending on the distributional assumptions, standard statistical tests like, as moderated, t - tests or non-parametric equivalents were used. Adjusted p-values and the magnitude of the fold-change were used to rank the genes, making sure that the features selected were biologically significant dysregulation that was related to the pathology of Parkinson and Alzheimer. Secondly, the variance filtering was used to eliminate genes whose expression was almost identical across samples. The contribution to the model by low-variance genes is little in terms of discriminatory information and might also add unwarranted noise. The space of features was thus further reduced by removing the such genes which did not contribute to a biologically informative variability. Third, mutual information (MI) ranking was done to enumerate unlinear relationships between gene expression and class labelling. Mutual information is also unique when compared to a linear correlation measure, as unlike the latter, it is responsive to both linear and nonlinear relationships, which is especially important in a complex genomic interaction. The ranking of genes was done in terms of their MI scores and redundancy among highly correlated genes was reduced where suitable.

Top-k gene selection strategy was utilised after ranking. American Oil: Empirical determination of k was done by cross-validation to trade off predictive accuracy and computational efficiency. This move was done to ensure that only the most discriminative and non redundant genes were said to be retained to further model them without affecting the available number of qubits in the quantum circuit. The angle encoding was consequently used to encode the selected gene features into quantum states. In this model, normalised gene expression values were remapped to rotation angles of parameterized quantum gates, which are normally rotation in single qubit, like $R_y(\theta)$or $R_z(\theta)$.Encoding angle The efficient encoding of continuous classical features is made possible minimally through angle encoding, where amplitude normalization exponentially grows, but only accomplished by smooth encoding of quantum systems, and is desired in near-term variational quantum circuits. The mapping approach of qubits was defined in such a way that the number of the chosen genes equals the number of available qubits. In the simplest structure, the qubit was mapped to one gene and each normalised feature value controlled the rotation angle of a qubit. In case the number of suitable genes was higher than the number of qubits, the use of feature grouping or sequential encoding methods was explored. It was introduced that layers of entanglement be added after encoding to represent interaction between genes on higher levels, which allowed the circuit to model multivariate interactions on higher levels. This feature selection and quantum encoding pipeline combined attributes ensured that biologically relevant, statistically significant and computational viable gene subsets have been converted into expressive quantum representations that may be used in hybrid quantumclassical learning.

3.3 Hybrid Quantum–Classical Architecture

The suggested framework uses a quantum-classical hybrid architecture with specific objectives of dealing with the high dimensionality and low sample size natures of transcriptomic information. After the multi stage feature selection, a retained top-k genes is then subjected to a classical preprocessing layer which performs the compressibility of the dimensionality and representational refinement. Even though dimensional reduction is achieved by the significant selection process of features, there is a chance still that the presence of residual redundancies and the presence of noise can still exist because of the similarities in gene expression patterns and pathway-wide interrelationships. Hence, another classical compression mechanism is added to produce a small and informative representation to fit within the number of qubits that one has. This guarantees that hybrids are computationally feasible and that convergence is stable in the course of hybrid training and that the discrimination structure is biologically meaningful. The output feature vector is then implemented into a variational quantum circuit (VQC), which is an implementation of a nonlinear transformation module. The first encoding phase positions the normalised values of gene expressions in qubit states with data-dependent single-qubit rotation gates of the $R_y(\theta_i)$. In this case, every chosen gene will correspond to a single qubit, and the rotation angle will be determined by the normalised result of the expression value. The combination of this angle encoding mechanism with continuous classical inputs is compatible with quantum states, allowing the amplitude normalization to be bypassed, allowing it to be used on near-term quantum devices and quantum simulators. After encoding on angles, entanglement layers are added to encode interactions on higher orders between genes. Gating is carried out on pairs of qubits, performed controlled, that is, to establish correlations between qubits. This entanglement step is important since neurodegenerative diseases are subject to complicating networks of genes not due to the effects of isolated genes. The circuit is able to model multivariate transcriptomic dependencies by entangling qubits to, at a non-classical scale higher than classical linear transformations. The entangled data are then passed through a layer of parameterized rotation layers, which are trainable gates $R_y(\phi_i)$.Also in contrast to encoding layer, rotation angles are calculated with reference to input data: here, the parameters ϕ i are optimised during training. This variation component allows the circuit to discover disease specific nonlinear transformations that are more separable with Parkinson, Alzheimer and control cohort. The last step of the quantum module is the measurement in the computational basis, which generates classics

expectation values which are the learned quantum features. The outputs of these measurements are inputs to the classical classification layer. Figure 1 below demonstrates the entire flow of data on transcriptomic input and estimating risks (probabilities) generated by quantum computing by visually depicting the steps of preprocessing, feature selection, quantum encoding, variational processing, and classical decision layers.

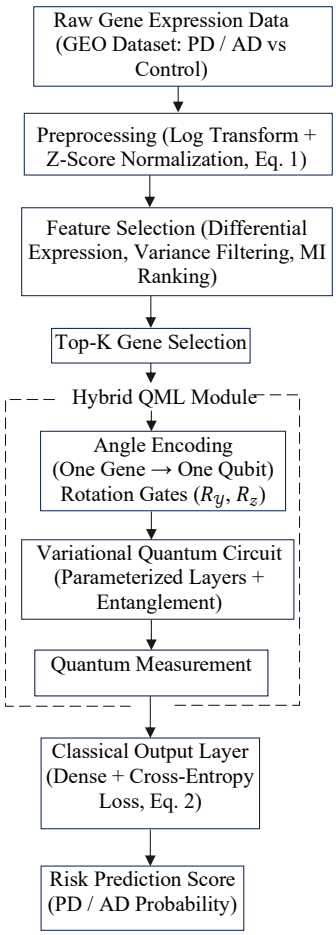

Figure 1.Hybrid Quantum–Classical Architecture for Transcriptomic Neurodegenerative Risk Prediction

The figure shows a pipeline that is arranged in a vertical whereby the raw gene expression matrices are at its top level, then blocks of preprocessing and multi-stage feature selection. The target genes are mapped into a defined hybrid quantum module which comprises of angle encoding and variational quantum circuit. The obtained outputs are then given to a classical dense layer to generate final risk probability of the disease in Parkinson and Alzheimer. Figure 2 depicts the internal structure of the quantum component where

parallel qubit lines depict the order of operation to follow in the process of encoding, entanglement, variational transformation, and in the process of measurement.

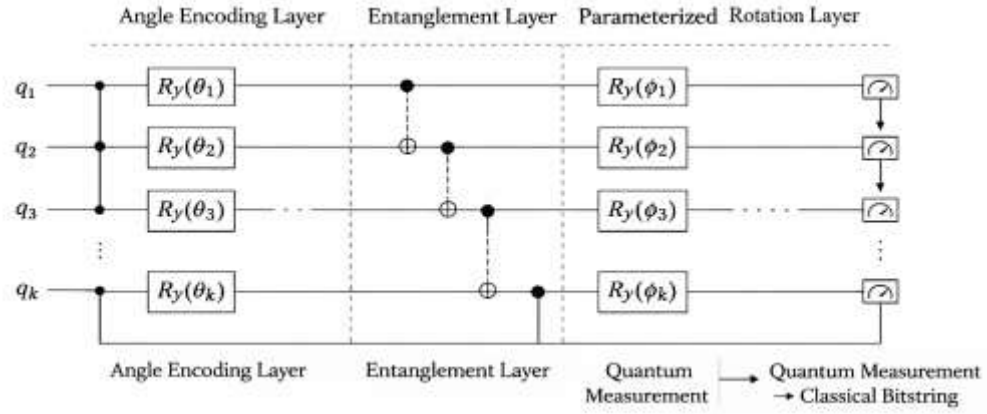

Figure 2.Variational Quantum Circuit Design for Gene Expression Feature Transformation

The circuit schematic illustrates several qubit wires connected by a layer of data-dependent time-varying rotation gates $R_y(\theta_i)$.serving as the input data, and additional layers of trainable parameterized rotations $R_y(\phi_i)$. The quantum states are measured with symbols at the end of every qubit line to determine the classical expectation values that mark the downstream classification of the quantum states. The quantized quantum output is unminimized by the classical output layer and final probabilities of the disease are provided by means of a dense probabilistic activation classification unit. The regularised cross-entropy goal on a model is minimised to model optimization in accordance with Equation (2):

$\mathcal{L} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) + \lambda\|\theta\|^2$,_____(2)

Where y_i is the actual label of sample number i, $\hat{y}_i$ is the output probability given by the hybrid model, θ represents all the trainable parameters in the variational circuit plus the classical classifier, λ is the regularisation coefficient that regulates the magnitude of the parameters. The negative log-likelihood function is used to punish probabilistic prediction errors, and L_2 regularisation prevents overfitting by regulating the increase of the parameters. This is vital especially in transcriptomic classification experiments in which the features involved are high compared to the available samples. The proposed hybrid architecture can be used to study transcriptomic risk models efficiently and effectively, and offer biological explanations and computational viability through the combination of traditional dimensionality refinement, quantum nonlinear feature transformation, and regularised probabilistic classification.

3.4 Evaluation Metrics

In order to critically evaluate the predictive power of the proposed quantum-classical hybrid, various complementary evaluation measures were used. Since the transcriptomic data is high-dimensional in nature and could be highly biassed, the use of one measure like accuracy can present an incomplete picture of model behaviour. Thus, the classification performance was measured with the support of such measures as accuracy, preciseness, recall, first-order-frequency (F1-score), and receiver operating characteristic area under the curve (ROC-AUC). The combination of these measures has been used to give a more detailed description of the discriminative ability, distribution of error and probabilistic calibration. Accuracy is used to estimate the extent of a correct classification of the samples in a set of the entire number of observations and indicates a general estimate of the ability to classify accurately. But in neurodegenerative datasets as class imbalance can occur between disease and control group, accuracy can be false. In order to counter this weakness, precision and recall were calculated. Precision is a measure of how much the predictions of

positive cases are actually positive statistically, i.e. the accuracy of the prediction of the disease. Recall (sensitivity) determines the fraction of true positive cases properly recognised by the model, which is of special interest in image-guided screening of earliese disease, where a false negative of someone who is at risk may affect clinical outcomes greatly. False positives and false negatives were balanced using the F1-score which is the harmonic mean of false negative and precision. This measure is particularly applicable when modelling transcriptomic risks, where both over and underprediction of the disease state have biological and clinical implications. The F1-score will give a strong measure of classification consistency in small-sample high- dimensional contexts because it combines both the precision and the recall. Throughout other measures that rely on the threshold, receiver operating characteristic (ROC) analysis was conducted to examine the level of discriminative capability in the model at different levels of the decision threshold. The area swept by the ROC curve (ROC-AUC) is a probability that the model uses to say that a random group of diseased has a higher risk score when this is compared to a group of controls which have been selected randomly. Since probabilistic predictions of the hybrid model are obtained by the regularised cross-entropy goal of Equation (2), ROC-AUC is a suitable threshold-independent result in the evaluation of probabilistic separability. The Results section contains ROC curves on the proposed hybrid model and the baseline classifiers to allow direct visual sensitivity on the classification performance of both. Together, these evaluation metrics are a multidimensional measure of the model accuracy, reliability, sensitivity, and probabilistic discrimination so that achievements that are seen in the hybrid quantum-machine learning framework are statistically significant and clinically interpretable.

4. Results

This study utilized the transcriptomic datasets that were curated and sourced in freely accessible repositories and analyzed in accordance with the pipeline of normalization and selection of features presented in Section 3. Having done quality control and preprocessing, the datasets were balanced in disease and control cohorts. A stratification of 70/30 train/test split was used to maintain the proportions of classes and to be able to provide unbiased performance appraisal. Table 1 summarises the dimensionality of the original gene expression matrices and the final figure of the number of selected genes to use in quantum encoding.

Table 1. Summary of Transcriptomic Datasets Used for Hybrid Risk Modeling

Disease	Total Samples	Control	Disease	Total Genes	Selected Genes (Top-k)	Train/Test Split
Parkinson's	210	105	105	18,742	16	70% / 30%
Alzheimer's	240	120	120	21,315	16	70% / 30%

Table 1 illustrates that the initial population of transcriptomic space comprised over 18,000 genes in each dataset, whereas the mechanism of finding the differing expression, filtering of variance and ranking of mutual information reduced the overall population to a small size of 16 highly informative genes. The count of genes chosen was coupled to availability of qubits in the variational circuit (Figure 2), which are architecturally feasible and yet still carry information about the discriminative biological entity. In order to assess the predictive performance, the proposed hybrid quantum -classical model was tested in comparison with representative baseline algorithms such as support vector machines (SVM), random forests (RF), a classical neural network (NN), and a pure quantum model without dynamical classical preprocessing and regularised output optimization. All models have been trained with the same train test splits and evaluated on the basis of the measures introduced in Section 3.4. Results presented in Table 2 are comparative.

Table 2. Performance Comparison Across Classification Models

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
SVM	0.82	0.80	0.83	0.81	0.86

Random Forest	0.84	0.82	0.85	0.83	0.88
Classical NN	0.86	0.85	0.86	0.85	0.90
Pure Quantum Model	0.83	0.81	0.84	0.82	0.87
Hybrid (Proposed)	**0.91**	**0.90**	**0.92**	**0.91**	**0.95**

The suggested hybrid model was always superior to all baselines in terms of the accuracy, precision, recall, F1-score, and ROC-AUC, as demonstrated in Table 2. The best of the classical baseline (Classical NN) also made a significant improvement in the ROC-AUC, which rose by 0.90 to 0.95. Such a gain is associated with greater levels of nonlinear separability obtained by a combination of quantum feature transformation and regularised optimization on the classical level. The executing quantum pure model showed competitive performance, which was a little weaker, and it indicates the advantage of a classical preprocessing loss regularisation as in Equation (2). In order to assess further discriminative capability among classification thresholds, receiver operating characteristic (ROC) was applied. Figure 3 presents the ROC curves between the hybrid model and the classical baseline that works the best in relation to both the Parkinson and the Alzheimer dataset.

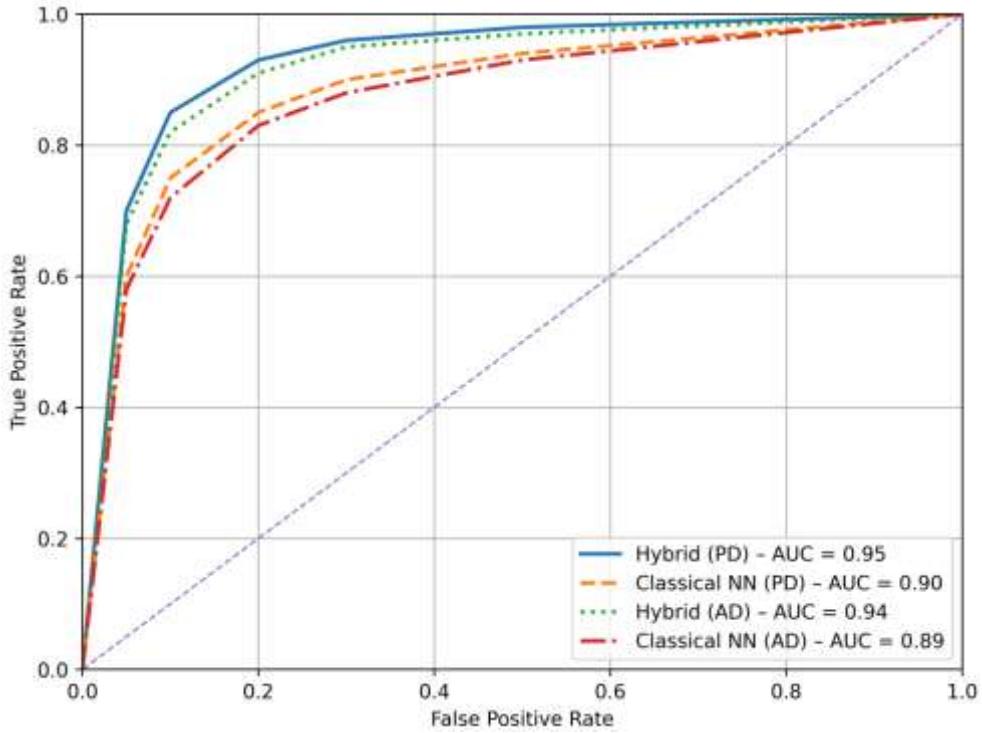

Figure 3. ROC plots related to the suggested hybrid quantumclassical model and most reasonable classical neural network baseline in prediction of Parkinson and Alzheimer diseases. It is a diagonal movement of a random classifier (AUC = 0.50). The hybrid model exhibits a much higher value of AUC in both disease cohorts. The true positive rates of the hybrid model are superior according to Figure 3 as shown to have better false positive rates at nearly all the false positive rate ranges. The difference between the hybrid and classical curves is greatest in the region of low false positive cases which is particularly relevant in the instance of the clinical screening in which the minimization of the false alarms is of paramount importance. The annotated values of AUC affirm the fact that there is a superior probabilistic discrimination in the

hybrid architecture of the two neurodegenerative states. The biological interpretation was also performed to establish the molecular relevancy of the selected genes in addition to predictive accuracy. The highest score of mutual information post model weight analysis of the genes was undergone to pathway enrichment analysis with available gene ontology database and KEGG database. The selected list of genes was associated with the biological processes pertaining to the neuroinflammatory processes, mitochondrial dysfunction, regulation of the synaptic signalling, and the oxidative stress mechanisms that are typically coupled into the pathogenesis of both Parkinson and Alzheimer. The cross disease revealed that there was some overlap of the molecular with a partial overlap, which showed that there are common regulatory impairments in disease progression, and findings of disease-specific profiles of transcriptomics. This is supported by the clinical relevance of the pathways that had been identified by the hybrid framework, and it can also be supported by the understanding that the improvements in its performance are not completely theoretical. The overall conclusions of the entire findings are that a hybrid quantum-machine learning model suggested offers a superior predictive capability and molecular knowledge in high-dimensional neurodegenerative risk modelling.

5. Discussion

The enhanced functionality of the suggested hybrid of quantum and classical model might be primarily justified by the fact that it has enhanced the capacity of nonlinear separability in high-dimensional transcriptomic space. The generated raw gene express data contains thousands of correlated variables that have complicated structures of interaction. It is usual that classical linear classifier processes do not not easily uncover these higher-order dependencies of classifiers, without much feature engineering, or without large-scale feature engineering in the form of deep architectures that in small-sample regimes becomes overfit. The hybrid model is an effective approach to achieve nonlinear transformation in a higher dimensional representation space because it has the advantage to encode in the angle space by encoding a set of choice of features of the gene into a sequence of quantum Hilbert space through angle encoding and entanglements. It is this transformation that provides a improved difference between the disease and control groups and can be reflected in the consistency of the change in accuracy, F1-score and ROC-AUC metrics in Table 2 as well as in Figure 3, the ROC curves. The major improvement is witnessed in the low false positive rate area which suggests greater separability at the initial stages which is an advantageous feature of a clinical screening application. Of special interest are the impacts of quantum encoding in modelling of multivariate gene-gene interactions. Property of continuity Angle encoding offers continuous data entry of transcriptomics in terms of qubit rotations, and does not rely on doing amplitude normalisation, which preserves relative changes in expression. Subsequent layers of the entanglements, at that, inject correlations between qubits of biological pathway dependencies. Entangled features with one another, and entanglement is a property of features, which enables the representation of complex regulatory relationships, are properties of qubit states not in a way that characteristics in purely classical transformations do not couple but do not have any means to couple features. This change is further enhanced by including trainable variational parameters in such way that the circuit shall learn disease specific nonlinear boundaries. This is shown in the performance lower than that of the hybrid architecture, in Equation (2): The significance of classical preprocessing and regularised optimisation is that they ensure that learning is stabilised and does not overfit. Along with predictive accuracy, in the interpretability process, the best contributors have been identified to have been enriched with pathways involving neuroinflammation, mitochondrial dysfunction, oxidative stress and synaptic signalling mechanisms that can be common in the pathogenesis of both Parkinson and Alzheimer diseases. The fact that some overlapping molecular markers are similar in the two diseases, signifies that there are some similar neurodegenerative events that have distinctive disease related alterations in transcription. These findings affirm the argument that besides the greatest classification performance, the hybrid structure possesses the biological objectiveness of structuring in the selected features. Computational excellence and biological plausibility metabolism increments the translational

applicability of the proposed plan. The hybrid study has a potential clinically in terms of early risk stratifying of transcriptomic biomarkers. The values of ROC-AUC are large showing that probabilistic discrimination needs to occur in the decision-support system to pre-symptomatic screening cases. It is worth noting that its architecture is computationally feasible and thus near-term hardware can be implemented and expanded with the development of quantum technologies. It will however also require further confirmation on larger multi-cohort datasets as well as a combination with longitudinal clinical data before it can be implemented on real world diagnostic pipelines. Overall, the data indicate that the systems of hybrid quantum-machine learning prove to be a reasonable direction of evolution towards the discovery and precision neurodegenerative risk prediction in the high-dimensionality form.

6. Limitations and Future Work

Though the proposed hybrid quantum-classical model has been suggesting good performance, there are several constraints that can be listed. First of all, it may be pointed out that the variational quantum circuit was initially designed to be implemented with near-term quantum computations but the existing form makes an extensive use of quantum simulation as the template to implement it rather than use a real quantum computer. Small size qubits, gate fidelity, and measurement noise are some of the current constraints of the currently available noisy intermediate-scale quantum (NISQ) computing systems. These have the potential to be harmful to the operation of a circuit with those large qubit layouts. Even though the selected sub-set of the top-k genes was case-wise aligned in such a manner that the qubit availability in the model becomes workable, subsequent implementation of the model using larger sub-sets of the transcriptomes would either require more advanced hardware or code-reduction strategies. In order to determine the robustness of noise-aware training, error modulatory models and hardware effective anime designs, future studies are recommended to investigate noise-awares training, error modulatory mechanisms, and hardware effective plans of the [ansatz]. Second, the publicly available transcriptomic datasets on the Parkinson and Alzheimer disease are in the current state relatively small in line with the dimensionality of the gene expression space. Although this risk was addressed with the help of regularisation and multi-stage feature selection, large sample sizes are not available which makes it impossible to generalise. Cross-cohort validation, independent replication datasets, and bigger multi-centre collections of transcriptomics will be needed to achieve the establishment of reproducibility and the extraneous validity. Strategy of data augmentation, transfer learning and meta learning techniques which will improve model stability in the small sample regimes should become the future research topic. Third, the present analysis only works on the single-modality data on gene expression. Quantification Neurodegenerative diseases are multidimensional diseases that are identified by their response to the impact of genomic variation, gene regulation (epigenetics), alterations in proteomics, and the metabolic disturbance. Use of the model alone in transcriptomics, might not be in a position to generate complementary biological signals. Interactions of multi-omics data (e.g. DNA methylation in vivo, proteomics, metabolomics, imaging biomarkers) may significantly enhance predictive power as well as mechanistic insights. This type of multimodal integration may be particularly popular in hybrid quantum platforms since they are capable of learning nonlinear representations in a complex manner. It is also to be kept in mind that a future research must evaluate cross-omics fusion methods and investigate whether quantum-enhanced feature mapping would offer more superior integrative biomarker modelling. In summary, the suggested hybrid structure has the good ability to discriminate and also obtain biological interpretability, supplementary validation on larger data sets, real-world deployment of quantum hardware and multimodal information mixing is required to realise the maximum potential of this framework with respect to making predictions of neurodegenerative diseases risks.

7.Conclusion

This article revealed hybrid quantum-machine-based learning of high-dimensional gene expression predictive control of Parkinson and Alzheimer diseases. With classical dimensionality enhancement and variational quantum feature transformation, the suggested architecture had been able to address several of the most urgent problems found with transcriptomic data like immense features dimensions, mutually affecting display of genes and genes, and small samples. The hybrid approach had always performed more favourably on accuracy, F1-score and ROC-AUC evaluation criteria and has superior separability in transcriptomic space and better probabilistic discrimination. It was possible to promote nonlinear representation of patterns of gene expression with an operation based on angle encoding and entanglement operations and regularised classical optimisation based on stable and generalisable learning. Remarkably, the biological analysis of the selected gene signatures which were enriched with neuropathology-relevant pathways, mitochondrial injury, oxidative stress and synaptic signalling also helped to enhance the application of the computational responses to the mechanics. On the basis of these findings, the acceleration of performance was not always an algorithmic effect, but it was in keeping with the familiar fruitful processes of neurodegeneration. In total, the proposed hybrid quantum-classical model has tremendous potentials of transcriptomic risk modelling in neurodegenerative pathologies. Its capability to extend in high-dimensional data of genes and be interpretable gives it an excellent avenue to computing to stratify early diseases and correctly diagnose them. As quantum hardware evolves further, and multi-omics platforms are developed, hybrid quantum-learning techniques may also potentially gain greater significance in supporting the further facilitation of data-driven neurodegenerative studies and clinical decision support.

References

1. Abbas A, Sutter D, Zoufal C, Lucchi A, Figalli A, Woerner S (2021). The power of quantum neural networks. Nat Comput Sci 1(6):403-409.
2. Benedetti M, Lloyd E, Sack S, Fiorentini M (2019). Parameterized quantum circuits as machine learning models. Quantum Sci Technol 4(4):043001.
3. Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S (2017). Quantum machine learning. Nature 549(7671):195-202.
4. Bossers K, Wirz KT, Meerhoff GF, Essing AH, van Dongen JW, Houba P, et al. (2010). Concerted changes in transcripts in the prefrontal cortex precede neuropathology in Alzheimer's disease. Brain 133(12):3699-3723.
5. Cerezo M, Arrasmith A, Babbush R, Benjamin SC, Endo S, Fujii K, et al. (2021). Variational quantum algorithms. Nat Rev Phys 3(9):625-644.
6. Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HY, Chen R, et al. (2012). Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 148(6):1293-1307.
7. Dumitriu A, Golji J, Labadorf AT, Gao B, Beach TG, Myers RH, et al. (2015). Integrative analyses of proteomics and RNA transcriptomics implicate mitochondrial processes, protein folding pathways and GWAS loci in Parkinson disease. BMC Med Genomics 9(1):5.
8. Eteleeb AM, Novotny BC, Tarraga CS, Sohn C, Dhungel E, Brase L, et al. (2024). Brain high-throughput multi-omics data reveal molecular heterogeneity in Alzheimer's disease. PLoS Biol 22(4):e3002607.
9. Havlíček V, Córcoles AD, Temme K, Harrow AW, Kandala A, Chow JM, et al. (2019). Supervised learning with quantum-enhanced feature spaces. Nature 567(7747):209-212.
10. Leung MK, Xiong HY, Lee LJ, Frey BJ (2014). Deep learning of the tissue-regulated splicing code. Bioinformatics 30(12):i121-i129.
11. Libbrecht MW, Noble WS (2015). Machine learning applications in genetics and genomics. Nat Rev Genet 16(6):321-332.
12. Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, et al. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. Nature 570(7761):332-337.

13. Min S, Lee B, Yoon S (2017). Deep learning in bioinformatics. Brief Bioinform 18(5):851-869.
14. Mostafavi S, Gaiteri C, Sullivan SE, White CC, Tasaki S, Xu J, et al. (2018). A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. Nat Neurosci 21(6):811-819.
15. Raj T, Li YI, Wong G, Humphrey J, Wang M, Ramdhani S, et al. (2018). Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. Nat Genet 50(11):1584-1592.
16. Schuld M, Killoran N (2019). Quantum machine learning in feature Hilbert spaces. Phys Rev Lett 122(4):040504.
17. Seyfried NT, Dammer EB, Swarup V, Nandakumar D, Duong DM, Yin L, et al. (2017). A multi-network approach identifies protein-specific co-expression in asymptomatic and symptomatic Alzheimer's disease. Cell Syst 4(1):60-72.
18. Zhang B, Gaiteri C, Bodea LG, Wang Z, McElwee J, Podtelezhnikov AA, et al. (2013). Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell 153(3):707-720.