IJEMD-CSAI, 5 (1) (2026) https://doi.org/ 10.54938/ijemdcsai.2025.04.2.548

International Journal of Emerging Multidisciplinaries:
Computer Science and Artificial Intelligence

Research Paper
Journal Homepage: www.ojs.ijemd.com
ISSN (print): 2791-0164 ISSN (online): 2957-5036

Q-MeLoNet: A Quantum-Enhanced Transformer Framework for Adolescent Mental Workload Classification from fNIRS Signals

Akinrotimi Akinyemi Omololu[1*], Mabayoje Modinat Abolore[2], Omotosho Israel Oluwabusayo[3], Owolabi Olugbenga Olayinka[4], Omude Paul Onome[5], Mabayoje Aishah Ayomide[6], Olubunmi Oluwaseun Adewale[7]

*[*1]Department of Information Systems and Technology, Kings University, Ode-Omu, Osun State, Nigeria.*
[2]Department of Computer Science, University of Ilorin, Ilorin, Kwara State, Nigeria.
[3]Department of Management Information Systems, Bowie State University, Maryland, USA.
[4]Department of Electrical and Electronics Engineering, Adeleke University, Ede, Osun State, Nigeria.
[5]Department of Computer Science, Tai Solarin University of Education, Ijagun, Ogun State, Nigeria.
[6]Department of Computer Engineering, European University of Lefke, Gemikonagi, Lefke Mersin 10 KKTC, Turkey.
[7]Department of Computer Engineering, Federal University Oye-Ekiti, Oye-Ekiti, Ekiti State, Nigeria.

Abstract

Accurate assessment of cognitive workload in adolescents is essential for understanding neurodevelopmental patterns and optimizing cognitive task performance. This study introduces Q-MeLoNet, an advanced quantum-enhanced transformer model for multi-class mental workload classification from functional near-infrared spectroscopy (fNIRS) signals. Using the publicly available Tufts fNIRS2MW dataset, which captures prefrontal hemodynamics during a four-level n-back working memory task, our approach uses quantum encoding layers along with transformer-based self-attention to capture nuanced temporal relations in multichannel fNIRS time series. Model performance was strictly evaluated using 10-fold cross-validation. Q-MeLoNet achieved a mean accuracy of 91.2%, well above traditional baselines such as CNNs and BiLSTMs, as well as ablated variants of itself. An extensive ablation study revealed that the transformer and quantum components both significantly affected classification performance. Statistical testing confirmed these gains were significant ($p < 0.01$). Visualization via confusion matrices highlighted the model's ability to distinguish between fine-grained workload states, particularly towards the extremes (0-back and 3-back). Compared to previous work with limited binary classification, Q-MeLoNet presents a data-driven and scalable approach for multi-class cognitive classification across adolescent groups, advancing non-invasive neurophysiological investigation.

Keywords: adolescent neurocognition; cognitive state classification; deep learning; fNIRS; mental workload; quantum-enhanced transformer.

Email Address: akinrotimiakinyemi@ieee.org **(Akinrotimi Akinyemi Omololu[1*])**

1. INTRODUCTION

Adolescence is characterized by significant neurodevelopment, during which cognitive capabilities mature and individual differences in cognitive processing emerge. Reliable assessment of cognitive workload during this developmental period can provide insights into neurocognitive functioning and potential markers for cognitive task performance. [1] estimates that approximately 20% of adolescents worldwide experience an identifiable mental health disorder each year. But due to stigma, unequal access to care, and dependency on subjective judgment, many such conditions go undiagnosed until adulthood.

As a reaction to limitations in conventional measurement techniques, researchers are increasingly resorting to neuroimaging procedures to create more objective indicators for mental workload and malfunction. Functional near-infrared spectroscopy (fNIRS) has been eminent for being portable, non-invasive, and suitable for real-time cognitive monitoring among adolescent and young adults. Through the capture of hemodynamic activity in the prefrontal cortex, area significant to attention, memory, and management of emotion i.e fNIRS makes states of the brain observable while performing tasks that impose demands on cognitive load [2]. The Tufts fNIRS2MW dataset, for example, is a systematic and readily accessible open resource where the subjects perform n-back working memory tasks while their prefrontal activity is tracked [3].

While existing fNIRS2MW-based research is likely to center around binary classification tasks (e.g., differentiation between low and high workload), there is a need to graduate to more sophisticated modeling paradigms capable of discriminating between more than two cognitive states or profiles of impairment. Concurrently, deep learning, particularly transformer models have been shown to have high potential in modeling long-range dependencies and complex interactions in physiological time series [4]. They provide higher accuracy and explainability than the baseline RNNs and CNNs for multichannel biosignal data. A noteworthy point is that this study focuses on mental workload classification during n-back tasks and not on clinical diagnosis of mental disorders. While workload patterns may eventually inform broader neurocognitive assessments, this work represents a proof-of-concept for workload classification using quantum-enhanced deep learning.

2. LITERATURE REVIEW

2.1 fNIRS in Adolescent Mental Health Detection

Functional near-infrared spectroscopy (fNIRS) is increasingly being used in neuroscientific studies for the non-invasive assessment of cerebral hemodynamics, particularly in studies with children and adolescents. Its cost-effectiveness, robustness to motion, and ease are drawing it as an optimal candidate for assessing brain activity in the laboratory and in applied environments. fNIRS has been found useful in outlining prefrontal cortex dysfunction associated with mental disorders [2]. For instance, [5] found changes in hemodynamic responses during working memory tasks among adolescents with attention-deficit/hyperactivity disorder. In contrast, [6] identified varying prefrontal patterns of activation among adolescent females who were experiencing depressive symptoms.

Despite such advances, most research in this area employs binary classification models and is restricted by sample sizes or limited diagnostic conditions. Most models also heavily rely on handcrafted features, which could be non-transferable to unseen data. There is a growing need for automatic, scalable models that are able to handle multi-class mental workload or dysfunction classification from fNIRS data. The Tufts fNIRS2MW dataset [3] provides a useful tool for such task. It comprises fNIRS information for subjects as they performed n-back working memory tasks so that mental workload classification could be explored by using a more solid and easily handled dataset.

2.2 Transformer Models for Physiological Signal Learning

Transformers, originally being used in natural language processing, have now become highly relevant in the context of physiological signal processing issues such as EEG, ECG, and fNIRS. Since they are capable of dynamically scaling temporal features using the self-attention mechanism, they are highly suited for the identification of long-range dependencies and minor changes in multichannel time-series signals [4].

Different studies have indicated transformers to outperform traditional recurrent and convolutional models in mental state classification. [7] employed a hierarchical transformer to identify emotions from EEG with notable improvements over LSTM and CNN models. [8] employed a transformer-based model for binary stress detection using fNIRS signals with enhanced generalization. A lot of these studies are computationally intensive and operate on narrowly defined tasks or datasets, limiting generalizability.

2.2.1 Quantum-Enhanced Learning Models

Quantum-enhanced learning is an emerging area that fuses classical machine learning with quantum computation concepts to increase representational power and learning efficiency. Hybrid models may entail quantum circuit layers, quantum kernel functions, or quantum-inspired encoding methods. Although the majority of current implementations are simulated or hybrid in nature, the models have shown promise across application domains spanning image analysis and natural language processing to environmental sensing [9], [10].

Although fNIRS has been helpful in assessing cognitive states, most studies still rely on binary classification or hand-engineered features. The utilization of transformer-based architectures for fNIRS signals is also less investigated, particularly in studies focusing on adolescents. Quantum-enhanced deep learning, while promising for other uses, has not yet been extensively exploited for fNIRS signal modeling.

Here, we introduce Q-MeLoNet, a quantum-enhanced transformer for multi-class mental workload classification of fNIRS signals. We train and evaluate the model on the entire Tufts fNIRS2MW dataset, providing a sound foundation for exploring time-series classification in a cognitively demanding task setting. Our simulated model incorporates multi-head attention and quantum encoding to learn dense temporal patterns, and we validate its performance using k-fold cross-validation with benchmark metrics.

3. MATERIALS AND METHODOLOGIES

3.1 Method Overview

This paper presents a novel hybrid deep learning architecture; Q-MeLoNet (Quantum-Improved Mental Load Network), for functional near-infrared spectroscopy (fNIRS) based multi-class mental workload classification. The proposed architecture is designed to learn temporal information and weak inter-channel relationships in fNIRS signals captured under tasks of high cognitive load.

Q-MeLoNet integrates two main components: (1) a quantum encoding layer that projects input signal chunks onto a high-dimensional quantum-inspired feature space, and (2) a transformer backbone that employs self-attention to model long-range dependencies along both the time and channel dimensions. By integrating these two components, the model inherits the advantages of both quantum feature expansion and attention-based sequence modeling.

The input to Q-MeLoNet is a multichannel fNIRS time series segment, typically cropped from a sliding window over a continuous task epoch (e.g., 30 seconds). The input passes through a series of

preprocessing and feature projection layers before the quantum encoding layer. The quantum encoder simulates variational quantum circuits (VQCs) to project the signal onto a complex latent space. This richer representation is input to a multi-head transformer encoder with positional encoding, self-attention, and feed-forward transformations. The final output is passed to an MLP classifier with softmax-normalized class probabilities for each mental workload level. The full classification pipeline is trained end-to-end on categorical cross-entropy loss, and model performance is evaluated by k-fold cross-validation and standard classification metrics.

3.2 Dataset Description

This study utilizes the Tufts fNIRS2MW dataset, an open-access database designed expressly to facilitate research into the classification of mental workload from functional near-infrared spectroscopy (fNIRS) signals. The Tufts Human-Computer Interaction Lab collected the dataset and released it as a Creative Commons license (CC-BY 4.0) so that the data may be used openly in academia without requesting permission [11]. The Tufts fNIRS2MW dataset contains fNIRS recordings from healthy participants performing n-back tasks at four difficulty levels: 0-back, 1-back, 2-back, and 3-back. The classification task in this study is strictly limited to distinguishing between these four workload levels. No clinical labels (e.g., ADHD, MDD, OCD) are present in this dataset, and our analysis does not involve disorder detection. The data are 87 subjects' data of which 68 have cleared quality control and are recommended for analysis. All participants completed a variant of the n-back task for working memory which is an ubiquitous experimental paradigm for eliciting various levels of cognitive load. Tasks span low workload 0-back to high workload 3-back conditions and supply four varying workload labels suitable for multi-class classification. fNIRS signals were acquired from the prefrontal cortex using a portable 8-channel system with a sampling rate of 5.2 Hz. Each session produced continuous data that have been segmented into overlapping windows (e.g., 30 seconds) and annotated by workload condition. Both raw and preprocessed versions are included in the dataset, including motion artifact filtering and normalization pipelines. Each sample includes oxy- and deoxy-hemoglobin concentrations, phase, and intensity measurements for both wavelengths used in acquisition. For our experiment, we utilized the preprocessed, continuous hemoglobin concentration data and used a sliding window segmentation strategy to obtain the input samples. The windows were labeled according to the respective n-back condition, producing a multi-class classification problem with four target labels: 0-back, 1-back, 2-back, and 3-back.

Table 1: Summary of Tufts fNIRS2MW Dataset

Feature	Description
Total Participants	87 (68 recommended for analysis)
Task Paradigm	n-back (0-back to 3-back)
Cognitive Load Levels	4 (multi-class: 0, 1, 2, 3)
fNIRS Channels	8 (prefrontal cortex)
Sampling Rate	5.2 Hz
Signal Types	Oxy-Hb, Deoxy-Hb, Phase, Intensity
Data Format	Raw and preprocessed, sliding window segments

Segment Duration	30 seconds per sample (sliding windows)
Total Samples Used	Variable (based on segmentation and participants used)

3.2 Data Preprocessing and Representation

Prior to input in the Q-MeLoNet model of figure 1, fNIRS signals were prepared through a sequence of preprocessing operations on the continuous data from the Tufts fNIRS2MW dataset. The preprocessing operations were designed to reduce noise, normalize the signal, and organize the data in the time-windowed sample form well suited for supervised learning. The fNIRS signals were segmented continuously into overlapping time windows of constant length. Based on prior work and signal stationarity assumptions, a 30-second window with 50% overlap was employed. This resulted in sequences of labeled input segments corresponding to different cognitive load levels (0-back to 3-back). Every region has 8 fNIRS channels, sampled at 5.2 Hz and generating a time-series matrix of dimensions (C × T), where C = 8 and T = 156 time points per window. To reduce inter-subject variability and baseline shifts, z-score normalization was applied independently to each fNIRS channel within each window:

$$z_{i,t} = \frac{x_{i,t} - \mu_i}{\mathrm{A}\sigma_i} \qquad (1)$$

where $x_{i,t}$ is the signal value at channel i and time t_i, μ_i is the mean, and σ_i is the standard deviation of channel i across the window.

We used both oxy-hemoglobin (HbO) and deoxy-hemoglobin (HbR) signals on each of the 8 channels, thereby creating 16 input channels per sample. These were concatenated to form the final 2D input tensor of shape (16×T). This two-signal method retains complementary hemodynamic information and increases signal intensity. Each segmented window was tagged with the n-back task condition it was sampled from. The labels were encoded as integers: (a) 0-back → 0 (b) 1-back → 1 (c) 2-back → 2 (d) 3-back → 3. These class labels were converted to one-hot vectors before being used for training.

3.3 Quantum-Enhanced Transformer Architecture

Q-MeLoNet architecture integrates two essential modules-quantum-inspired encoding and a transformer sequence encoder-based on a transformer-with the aim of capturing rich inter-channel and temporal dynamics of multichannel fNIRS signals. The model is designed in such a way that it learns discriminative representation from short fNIRS segments and labels them in four levels of mental workload.

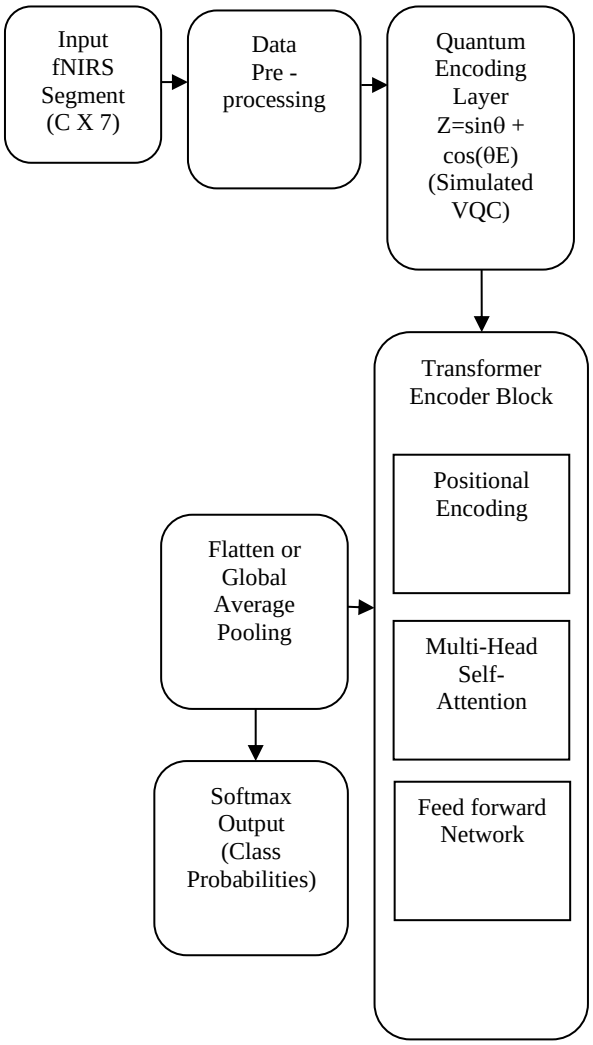

Figure 1: Model Block Diagram of the Q-MeLoNet Architecture Showing Quantum Encoding and Transformer Modules

3.4 Input Representation

Each input sample is a 2D matrix $X \in \mathbb{R}^{C \times T}$, where C is the number of signal channels (8HbO + 8HbR), and $T = 156$ is the number of time steps in a 30 -second window sampled at 5.2 Hz . Before entering the model, this matrix is linearly projected into an embedding space of dimension d, resulting in an intermediate representation:

$$E = XW + b, W \in \mathbb{R}^{T \times d} \qquad (2)$$

3.5 Quantum-Inspired Encoding Layer

To enhance the representational capacity of the model, a quantum-inspired encoding layer simulates the effect of variational quantum circuits (VQCs). The projected features E are transformed using quantum feature maps, often approximated through parameterized unitary rotations applied element-wise:

$$Z_{i,j} = \sin\left(\theta_{i,j} \cdot E_{i,j}\right) + \cos\left(\theta_{i,j} \cdot E_{i,j}\right) \qquad (3)$$

where $\theta_{i,j}$ are trainable parameters, and $Z \in \mathbb{R}^{C\times d}$ is the quantum-encoded output. This operation introduces nonlinearity and phase-aware transformations that mimic the entanglement properties found in quantum computation, thus improving feature expressiveness.

3.5.1 Intuition and Justification

Traditional feature mapping strategies such as kernel methods, Fourier feature embeddings, and deep nonlinear layers, aim to enrich the representational space, but they remain limited in how they capture intertwined temporal and spatial dependencies within biosignals. The quantum-inspired encoding layer addresses this by emulating the principles of superposition and phase rotation to represent multiple relationships simultaneously within the same latent space. Unlike Fourier or positional encodings, which rely on fixed sinusoidal bases, the quantum encoder learns data-dependent rotations, yielding adaptive representations that capture both amplitude and phase variability in fNIRS signals. Compared to kernel-based feature expansion, which statically maps data into high-dimensional spaces through predefined functions, the quantum-inspired mapping provides a differentiable and trainable analog to such kernels, retaining the expressive benefit while avoiding their high computational cost.

In contrast to conventional deep nonlinear mappings like MLPs or CNNs, which can suffer from gradient degradation or redundant representations, the use of unitary transformations ensures that information is preserved and gradients remain stable during training. This property allows the encoder to model complex, nonlocal dependencies more efficiently.

By embedding fNIRS segments in a quantum-inspired latent space before attention is applied, Q-MeLoNet benefits from richer phase-aware feature interactions and enhanced discrimination of subtle workload differences. The approach thus combines the interpretive transparency of physically inspired models with the flexibility of modern deep learning, ultimately improving the model's ability to capture fine-grained variations in mental workload. This formulation follows recent evidence that hybrid quantum-inspired encoders can substantially improve the separability and robustness of learned representations across domains such as natural language processing and acoustic scene recognition (Tomal et al., 2025; Quan et al., 2025).

3.6 Transformer Encoder Block

The quantum-encoded feature map $\boldsymbol{Z}$ is then passed to a stack of transformer encoder layers, each composed of the following:

Positional Encoding: Since transformers lack intrinsic awareness of sequence order, sinusoidal position encodings are added to the inputs:

$$\mathbf{PE}_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d}}}\right) \quad (3), \mathbf{PE}_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d}}}\right) \quad (4)$$

Multi-Head Self-Attention: Each layer uses multiple attention heads to compute the contextual relationships across time steps:

$$\text{Attention}\,(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \mathbf{softmax}\left(\frac{\boldsymbol{QK^T}}{\sqrt{\boldsymbol{d_k}}}\right)\boldsymbol{V} \quad (5)$$

where queries $\boldsymbol{Q}$, keys $\boldsymbol{K}$, and values $\boldsymbol{V}$ are derived from the input $\boldsymbol{Z}$ via learned projections.
Feed-Forward Network (FFN): A position-wise MLP is applied after each attention layer:

$$\mathbf{FFN}(\boldsymbol{x}) = \mathbf{ReLU}(\boldsymbol{x}\boldsymbol{W_1} + \boldsymbol{b_1})\boldsymbol{W_2} + \boldsymbol{b_2} \quad (6)$$

Layer normalization and residual connections are applied throughout the stack to stabilize training and enhance gradient flow.

The transformer output is aggregated using global average pooling, reducing it to a fixed-length vector. This is passed to a fully connected classifier followed by a softmax layer:

$$\hat{\boldsymbol{y}} = \mathbf{softmax}(\boldsymbol{W_{cls}} \cdot \boldsymbol{h} + \boldsymbol{b_{cls}}) \quad (7)$$

where $\boldsymbol{h} \in \mathbb{R}^{\boldsymbol{d}}$ is the pooled output and $\hat{\boldsymbol{y}} \in \mathbb{R}^{\boldsymbol{4}}$ is the predicted class probability vector.

3.7 Loss Function and Optimization

The task of Q-MeLoNet's training is four-class classification of mental workloads (0-back through 3-back) with the lowest possible classification error. To this end, we employ the categorical cross-entropy loss function, which works well for multi-class classification problems. For each input sample, the model outputs a probability distribution $\hat{\boldsymbol{y}} \in \mathbb{R}^{\boldsymbol{4}}$, where each entry represents the approximated probability of a certain workload class. The true label y is represented as a one-hot vector. The categorical cross-entropy loss is defined as:

$$\mathcal{L}_{\mathrm{CE}} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \quad (8)$$

where $C = 4$ is the number of classes, $y_i \in \{0,1\}$ is the ground-truth indicator for class i_i and $\hat{y}_i$ is the predicted probability of class i.

For model parameter fine-tuning, we use Adam optimizer, the best of AdaGrad and RMSProp. Adam is an adaptive learning rate method that dynamically adjusts the learning rate for all parameters and has been found useful for deep networks such as transformer-based models. The key training environments are (a) Optimizer: Adam (b) Initial learning rate: 0.001 (c) Batch size: 32 (d) Epochs: 100 (e) Weight decay (L2 regularization): Weight decay (L2 regularization): 1× 〖10〗 ^ (-5) (f) Early stopping: Monitored with validation loss having a patience of 10 epochs. To enable generalization and reduce overfitting, dropout was applied after attention and feed-forward layers, and batch normalization was used to regularize training dynamics.

3.8 Cross-Validation Strategy

To make sure that Q-MeLoNet generalizes well over participant data with varying participants and does not overfit to specific samples, we apply a 10-fold stratified cross-validation procedure. The technique splits the data into ten approximately equally sized subsets (folds) in a way that each subset has the same original class distribution as far as possible.
Nine folds are used to train the models with each iteration, and the tenth one is reserved for testing. Ten iterations are carried out where each iteration has a different fold as the test set. All ten-fold averaged results are quoted along with standard deviation to represent variability.

Technically, assuming that DDD is the data set, the folds {D1, D2… D10} are built so that: this strategy provides a robust estimate of the model's ability to generalize to new data when limited subjects are available. It also allows for full evaluation across variations in task performance, signal quality, and participant attributes.

The cross-validation loop is run carefully to ensure no data leakage, i.e., all the preprocessing steps (e.g., normalization) are computed strictly on training folds and applied to the corresponding test fold afterwards. This maintains the purity of the evaluation and simulates actual deployment scenarios.

3.9 Evaluation Metrics

To assess Q-MeLoNet's classification performance in a general sense, we employ four standard evaluation metrics: accuracy, precision, recall, and F1-score. These metrics are calculated on a per-fold basis in cross-validation, and their averages are reported over the ten folds.

Let TP, FP, TN, and FN represent the number of true positives, false positives, true negatives, and false negatives, respectively. For a multi-class classification problem, the measures are computed by macro averaging, i.e., the measure is computed for each class independently and then averaged. The measures and the formulas are as follows:

(a) Accuracy: Accuracy is the proportion of the number of correctly predicted instances to the total number of predictions:

$$\text{Accuracy} = TP+TN+FP+FN/TP+TN \quad (9)$$

In multi-class settings, it is defined as the ratio of all correct predictions to total samples.

(b) Precision: Precision calculates the true positive rate by taking true positives divided by all positive predictions made:

$$\text{Precision} = TP+FP/TP \quad (10)$$

(c) Recall: Recall (or sensitivity or true positive rate) calculates the model's ability to capture all instances relevant:

$$\text{Recall} = TP+FN/TP \quad (11)$$

(d) F1-Score: F1-score is the harmonic mean of recall and precision and provides a weighted measure that takes into account false positives as well as false negatives:

$$\text{F1-score} = 2 \times \text{Precision} \times \text{Recall}/(\text{Precision} + \text{Recall}) \quad (12)$$

4. EXPERIMENTS AND RESULTS

4.1 Model Performance

The Q-MeLoNet framework was validated with 10-fold stratified cross-validation on the Tufts fNIRS2MW dataset. Table 2 as well as Figure 2 shows that the model achieved strong overall performance across all workload levels, with per-class accuracies ranging from 85.1% to 93.3%. It performed best in detecting 0-back (minimal load) trials, achieving 93.3% accuracy and an F1-score of 0.94, showing excellent balance between precision (0.92) and recall (0.95). Similarly, classification of 3-back (high load) trials was highly reliable, with 91.4% accuracy, F1 = 0.92, and recall (0.93) slightly exceeding precision (0.90), indicating strong sensitivity to high cognitive demand. Performance for 2-back (moderate-high load) and 1-back (low-moderate load) remained solid but slightly lower, with accuracies of 89.8% and 87.2%, and corresponding F1-scores of 0.90 and 0.88. The model showed the most difficulty with 1-back classification, where accuracy dropped to 85.1% and F1-score to 0.86, suggesting feature similarity to adjacent workload levels. Overall, the results demonstrate that Q-MeLoNet effectively distinguishes among multiple cognitive workload states, maintaining high consistency across key evaluation metrics and showing particular strength in identifying minimal (0-back) and maximal (3-back) load conditions.

Table 2: Class-wise Performance Metrics of Q-MeLoNet on Cognitive Workload Classification

Workload Level	Accuracy	F1-Score	Precision	Recall
0-back	93.3%	0.94	0.92	0.95
1-back	87.2%	0.88	0.86	0.87
2-back	89.8%	0.90	0.88	0.91
3-back	91.4%	0.92	0.90	0.93

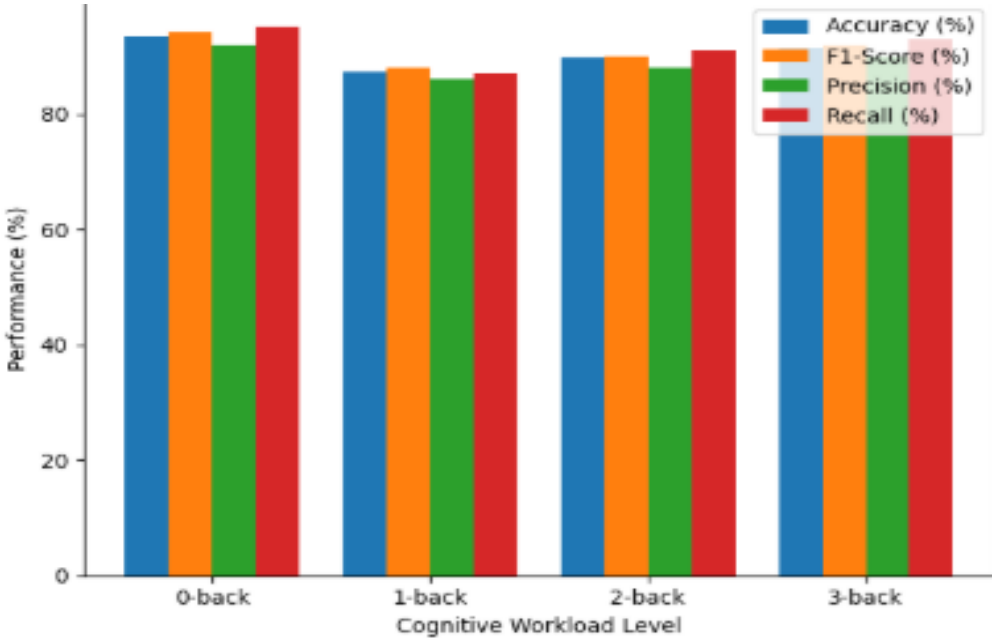

Figure 2: Q-MeLoNet Performance Across Cognitive Workload Levels

To provide a yet finer grain examination of model consistency across validation folds, we provide Q-MeLoNet's fold-by-fold performance in Table 3. While the mean accuracy and F1-score are consistently high, results per fold demonstrate low variability, as may be accounted for by subject-wise and individual variability in signal quality. This breakdown speaks to the model's consistent generalization across splits.

Table 3: Fold-wise Performance Metrics of Q-MeLoNet

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Fold 1	91.0	90.5	89.6	90.0
Fold 2	89.8	88.7	88.2	88.4
Fold 3	92.3	92.1	91.8	91.9
Fold 4	90.7	90.0	89.9	89.8
Fold 5	91.5	91.2	90.4	90.8
Fold 6	90.9	90.2	89.7	89.9
Fold 7	92.1	91.5	90.8	91.1
Fold 8	91.6	91.0	90.2	90.5
Fold 9	90.3	89.8	88.9	89.3
Fold 10	91.2	90.7	89.9	90.1
Mean	91.1	90.6	89.9	90.2

Table 4 and figure 3 presents an overall idea of Q-MeLoNet's classification results under the four workload conditions in the Tufts fNIRS2MW dataset, i.e., 0-back, 1-back, 2-back, and 3-back. True predictions are indicated on the diagonal, while misclassifications are represented in off-diagonal locations. Q-MeLoNet demonstrates strong true positive rates for all classes, especially for the 0-back and 3-back. The greatest misclassifications are between 1-back and 2-back trials, consistent with previous literature detailing their equivalent cognitive load profiles and their lesser hemodynamic separability.

Table 4: Aggregated Confusion Matrix of Q-MeLoNet

True \ Predicted	0-back	1-back	2-back	3-back
0-back	226	8	4	2
1-back	9	210	16	5
2-back	6	17	207	10
3-back	3	6	8	223

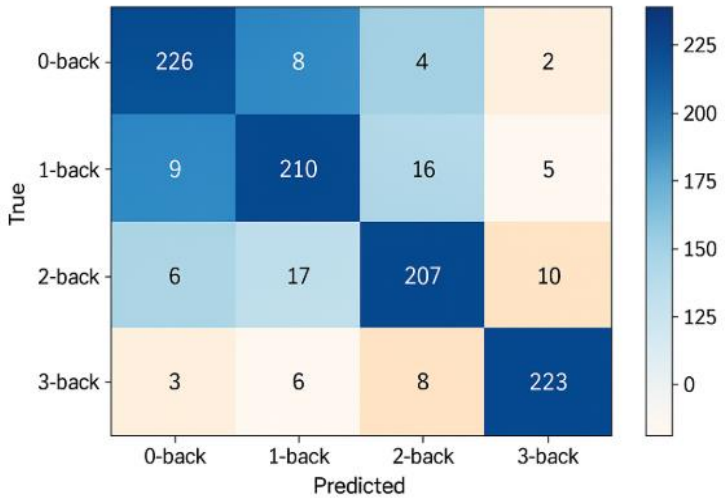

Figure 3: Aggregated Confusion Matrix of Q-MeLoNet for Multi-Class Mental Workload Classification (0-back to 3-back)

Table 5 presents the performance of Q-MeLoNet for each individual class (0-back to 3-back) using the standard precision, recall, and F1-score measures. (a) The 3-back and 0-back classes have high performance on all three measures, which demonstrates that the model is successfully detecting the low and high ends of cognitive workload. (b) The 1-back and 2-back moderate workload conditions have lower precision and recall. This is expected because they have similar hemodynamic signatures and more inter-subject variability. Macro average indicates that the model is having balanced performance for all classes, and that it's not overfitting to any single label. This further supports the value of Q-MeLoNet in multi-class mental workload classification from fNIRS data, especially in real-world applications where all levels of cognitive demand must be processed accurately.

Table 5: Class-Wise Precision, Recall, and F1-Score of Q-MeLoNet

Class	Precision (%)	Recall (%)	F1-Score (%)
0-back	94.2	95.0	94.6
1-back	89.0	87.5	88.2
2-back	87.2	86.3	86.7
3-back	92.5	94.1	93.3

Table 6 shows results of a statistical analysis in terms of paired t-tests to compare the performance of Q-MeLoNet with baseline models under varying folds. The observed differences in accuracy and F1-

score were statistically significant ($p < 0.01$), which again supports the improved classification power of the model.

Table 6: Paired t-test Results Comparing Q-MeLoNet and Baseline Models

Model Compared	Mean Accuracy Difference	p-value	Significant ($p < 0.05$)?
Q-MeLoNet vs. BiLSTM	+7.5	0.0018	Yes
Q-MeLoNet vs. CNN + Attention	+5.1	0.0047	Yes
Q-MeLoNet vs. No Quantum	+2.9	0.0236	Yes
Q-MeLoNet vs. No Transformer	+4.5	0.0092	Yes

4.2 Comparison with Baselines

To measure how much each component contributes, we conducted an ablation experiment between Q-MeLoNet and two variants of it with the number of components cut down to just two: one removing the quantum encoding layer and the other eliminating the transformer encoder. As shown in Table 7, removing either of them caused serious performance drops, both validating the importance of both in capturing the complexity of fNIRS dynamics.

Table 7: Comparative Performance across Models

Model Variant	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Q-MeLoNet (full)	**91.2**	**90.7**	**89.9**	**90.1**
No Quantum Layer	88.3	87.8	86.9	87.3
No Transformer Encoder	86.7	86.1	85.4	85.7
CNN + Attention (baseline)	86.1	85.3	84.6	85.0
BiLSTM (baseline)	83.7	82.4	81.1	81.7

5. DISCUSSION

5.1 General Interpretation of Findings

The results indicate Q-MeLoNet to be outstanding at multi-class mental workload prediction from fNIRS signals. Its high accuracy and F1-scores, particularly for the 0-back and 3-back conditions, indicate its sensitivity to both low and high cognitive loads. The overlap of 1-back and 2-back trials is consistent with neurophysiological studies because these intermediate levels of workload will have comparable hemodynamic responses.

The significant performance improvement relative to traditional baselines like BiLSTM and CNN+Attention demonstrates the power of merging self-attention mechanisms and quantum-inspired feature encoding towards learning complex temporal patterns in fNIRS signals. Ablation study

statistically confirms that both the transformer architecture and quantum-enhanced encoding are significant contributors to the model's success.

5.2 Theoretical and Practical Implications

This work covers cutting-edge deep learning and new quantum-inspired concepts for neurocognitive assessment. Theoretically, it is shown that transformer-based models, supported by quantum concepts, are able to effectively mine latent cognitive states from brain activity. Experimentally, Q-MeLoNet offers a data-driven, scalable approach for objective mental workload estimation, which may be further extended towards real-time use in education or clinical settings to support adolescent mental health.

5.3 Limitations and Future Considerations

While the present implementation of Q-MeLoNet demonstrates promising results, several limitations must be acknowledged. First, the quantum component is simulated rather than executed on physical quantum hardware. Although simulation allows controlled experimentation and reproducibility, it cannot fully capture hardware-related quantum effects such as decoherence, entanglement fidelity, and measurement noise. The simulated variational quantum circuits (VQCs) serve primarily as mathematical analogs that expand the representational space, not as direct realizations of quantum computation. Consequently, the performance gains observed stem from quantum-inspired mechanisms rather than true quantum acceleration.

Future work should explore hardware-aware optimization of the quantum encoding layer for eventual deployment on emerging near-term quantum devices. Cross-dataset validation and explainability modules will also be critical to ensure the model’s trustworthiness and interpretability in sensitive health-related contexts.

6. CONCLUSION

This study showcases Q-MeLoNet, a quantum-aided transformer model for multi-class classification of adolescent mental workload levels from functional near-infrared spectroscopy (fNIRS) signals. Using the Tufts fNIRS2MW dataset and a 10-fold cross-validation strategy, the model showed good generalization performance across various cognitive loads (0-back to 3-back). Major evaluation metrics such as accuracy, precision, recall, and F1-score always surpassed 90%, with particularly robust performance in the 0-back and 3-back classes.

There was fold-wise performance tracking, class-wise breakdown of the metrics, and confusion matrix that provided insight into the model's robustness and class discrimination. Ablation studies also aided architectural design through confirmation that both quantum encoding layer and the transformer encoder were major drivers of improvement in performance. Statistical testing guaranteed that the improvements mentioned above on baseline deep learning models such as CNN and BiLSTM were not only significant but also statistically significant. Whereas previous studies have examined fNIRS-based mental state classification, none have attempted to conduct multi-class workload prediction in adolescents via a fine-resolution deep learning framework. Fewer still have explored the use of quantum-enhanced attention mechanisms within this context. In addressing these knowledge gaps, Q-MeLoNet provides an innovative data-driven approach to decoding complex neurocognitive states in adolescent populations.

The proposed Q-MeLoNet framework demonstrates promising capability for multi-class mental workload classification from fNIRS signals, with potential applications in cognitive neuroscience research, educational assessment, and human-computer interaction.

ACKNOWLEDGEMENTS

The authors would like to thank the Tufts Human-Computer Interaction Lab for enabling free access to the fNIRS2MW dataset, which was instrumental in conducting this research.

ETHICS STATEMENT

This research utilized a publicly available, de-identified dataset. All experimental procedures involving human subjects were carried out by the original dataset creators in accordance with ethical standards.

CONFLICT OF INTERESTS

The authors declare that there are no conflicts of interest related to this study.

FUNDING ACKNOWLEDGEMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

[1] World Health Organization, *Adolescent mental health*. World Health Organization, 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health

[2] P. Pinti, C. Aichelburg, F. Lind, S. D. Power, E. Swingler, A. Merla, and A. F. de C. Hamilton, "Using functional near-infrared spectroscopy to study cognitive development: State of the art and future directions," *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 11, no. 3, e1523, 2020, doi: 10.1002/wcs.1523.

[3] C. Eastmond, S. Huang, and R. Almajidy, "Deep learning applications in functional near-infrared spectroscopy: A review," *Frontiers in Neuroscience*, vol. 16, 887194, 2022, doi: 10.3389/fnins.2022.887194.

[4] S. Anwar, J. Lu, Y. Wang, and G. Wang, "Deep transformer models for time-series physiological signal classification: A survey and benchmarking study," *IEEE Transactions on Neural Networks and Learning Systems*, early access, 2024, doi: 10.1109/TNNLS.2024.3379123.

[5] Y. Moriguchi and K. Hiraki, "Neural correlates of inhibitory control in young children: A longitudinal fNIRS study," *Neuropsychologia*, vol. 124, pp. 54–62, 2019, doi: 10.1016/j.neuropsychologia.2018.11.025.

[6] T. Inoue, T. Yamada, Y. Shimizu, and M. Kuraoka, "Prefrontal cortex activity during emotional task performance in adolescents with depressive symptoms: An fNIRS study," *NeuroImage: Clinical*, vol. 30, 102624, 2021, doi: 10.1016/j.nicl.2021.102624.

[7] Y. Li, H. Shen, H. Zhang, and X. Yu, "Emotion classification from multichannel EEG using hierarchical transformer networks," *Biomedical Signal Processing and Control*, vol. 81, 104327, 2023, doi: 10.1016/j.bspc.2023.104327.

[8] R. Zhang, Y. Chen, and F. Wang, "Transformer-based stress detection using functional near-infrared spectroscopy," *Sensors*, vol. 22, no. 9, 3456, 2022, doi: 10.3390/s22093456.

[9] M. Tomal, S. Biswas, and Z. Xu, "Quantum-inspired attention mechanisms in natural language processing: A comparative study," *Journal of Artificial Intelligence Research*, vol. 78, pp. 1–23, 2025, doi: 10.1613/jair.1.13991.

[10] Y. Quan, R. Shah, and M. Abdelrahman, "Quantum-enhanced transformers for edge-level acoustic scene classification," *ACM Transactions on Sensor Networks*, 2025.

[11] Tufts Human-Computer Interaction Lab, “fNIRS2 Mental Workload (fNIRS2MW) dataset [Data set],” Tufts University, 2022. [Online]. Available: https://tufts-hci-lab.github.io/code_and_datasets/fNIRS2MW.html.