Attention-based Quantum Transfer Learning and Transformers for Accurate Autism Detection in Children through Facial Image Analysis

Soham Bhoir1, Harshal Dave2, Karthikeyan S3,\* , Niveditha S4, Meenalosini Vimal Cruz5 , Sarveshwaran R6, Shreyanth S7

1KJ Somaiya College of Engineering, Mumbai, India, [soham.bhoir@somaiya.edu](mailto:soham.bhoir@somaiya.edu)

2KJ Somaiya College of Engineering, Mumbai, India, [harshal.dav](mailto:harshal.dave@somaiya.edu)[e@somaiya.edu](mailto:e@somaiya.edu)

3Dept of CSE(AIML), KPR Institute of Engineering and Technology, Coimbatore, India, [skarthi.sns@gmail.com](mailto:skarthi.sns@gmail.com)

4Department of Rajalakshmi Engineering College, Thandalam, Chennai, sniveditha0412@gmail.com

5Dept of IT, Georgia Southern University, USA,[mvimalcruz@georgiasouthern.edu](mailto:mvimalcruz@georgiasouthern.edu)

6Data Science and Engineering, Birla Institute of Technology and Science, Pilani,sarvesh9111@gmail.com

7Data Science and Engineering, Birla Institute of Technology and Science, Pilani, Rajasthan, India , shreyanth0810@gmail.com

***Corresponding Author:*** *skarthi.sns@gmail.com*

***Abstract*—This paper presents an innovative approach for automated autism spectrum disorder (ASD) detection in children using facial images. By combining attention mechanisms from transformer-based architectures with quantum transfer learning techniques, the proposed model aims to accurately classify children as either healthy or potentially autistic. The model extracts relevant features from facial images, capturing subtle nuances indicative of ASD, and incorporates quantum computing to enhance the learning process. Evaluation of a comprehensive dataset of 3,014 facial images demonstrates the superiority of the proposed model, achieving a classification accuracy of 88.23%. This outperforms traditional machine learning approaches and overcomes limitations associated with current ASD detection methods. The innovative approach offers a more accurate and efficient method for ASD detection, with potential applications in identifying other neurological and developmental disorders. The research evaluates the effectiveness of vision transformers and quantum transfer learning for automated autism spectrum disorder (ASD) detection in children. By examining the performance of these two technologies individually, this study provides insights into their suitability and efficiency for ASD detection. The findings contribute to earlier and more effective interventions, paving the way for the potential integration of vision transformers and quantum transfer learning in healthcare, particularly in pediatric neurology. This research holds promise for improving diagnosis and intervention outcomes for children with ASD.**

***Index Terms*—Quantum transfer learning, autism spectrum disorder, Quantum computing, Neurodevelopmental disorders, Facial features, Transfer learning, Transformers.**

1. INTRODUCTION

**H**

EALTHCARE is one of the most critical sectors that affect the quality of life and economic development of a country. The World Health Organization, known as the WHO, defines health as "a complete state of physical, mental, and social well-being and not merely the absence of disease or

infirmity." With advancements in technology, the healthcare industry has been significantly impacted. Researchers and healthcare practitioners are continuously exploring new technologies to improve patient care and provide better healthcare services.

Child healthcare, a critical aspect of society’s growth, has seen notable advancements due to artificial intelligence (AI) and machine learning (ML). Among the conditions under scrutiny, Autism Spectrum Disorder (ASD) stands out. ASD, affecting children’s communication and behavior, demands early diagnosis and intervention for better outcomes. Traditional ASD diagnosis relies on labor-intensive methods such as Autism Diagnostic Observation Schedule (ADOS), Autism Diagnostic Interview-Revised (ADI-R), and developmental history evaluation. These methods have limitations, including subjectivity and the potential for misdiagnosis. Hence, there is a growing need for more objective and efficient ASD detection methods.

Various research has investigated alternate ways of ASD identification in recent years, including the extraction of features [1], eye tracking [2], facial recognition [3]–[5], medical picture analysis [6], app development [7], and speech recognition [8]. Despite these advances, the demand for more reliable and efficient diagnostic procedures remains.

Therefore, in the field of ASD detection, there is a growing need for more efficient and reliable diagnostic methods. Recent advancements in deep learning have led researchers to explore the use of computer vision techniques for detecting ASD in children [9]. These computer vision-based methods analyze facial expressions and other visual cues to identify potential indicators of ASD and can be subjective and prone to errors. However, the accuracy of these methods is still

limited, and further research is required to improve their efficacy.

Researchers have investigated the use of quantum information processing in a variety of healthcare disciplines, including drug development, diagnostic imaging, and personalized medical care [10]–[12]. Quantum computing has the potential to revolutionize many fields of research and development, including healthcare. In contrast to traditional computers, which process data in bits (either 0 or 1), quantum computer systems use qubits, which may represent both 0 and 1 at the same time. This distinguishing feature allows for faster and more efficient processing of complicated algorithms.

Quantum computing has shown promise in a variety of sectors, including machine learning, because of its capacity to analyze massive volumes of data rapidly and reliably. Although quantum computing is still in its early phases of development in the context of ASD diagnosis, promising results have been found utilizing quantum-inspired algorithms. Quantum transfer learning, in particular, has emerged as an intriguing strategy for improving the accuracy of deep learning models for ASD detection [4].

Quantum transfer learning leverages pre-trained models on classical computers and transfers the acquired knowledge to a quantum computer, thereby improving the precision of neural network models for ASD identification. However, further research is necessary to validate the effectiveness of quantum transfer learning and to develop specialized hardware tailored for quantum computing in healthcare applications.

Moreover, alongside quantum transfer learning, another promising approach for ASD detection is the use of transformer-based models. Transformers have demonstrated exceptional performance in a wide range of tasks in natural language processing [13], [14] and computer vision [15], [16]. They excel in capturing global dependencies and learning intricate patterns, leading to significant advancements in these domains. Applying transformer-based architectures to facial image analysis can potentially raise the accuracy and resiliency of ASD detection models.

This paper presents a novel approach for the automated detection of ASD in children using facial images. The proposed method integrates quantum transfer learning and transformer-based models to create a comprehensive framework that captures both local and global facial features.

By leveraging the capabilities of quantum computing, the learning process is enhanced, enabling more accurate detection of ASD.This approach has been evaluated on a dataset of 5,000 children’s face photographs, including autistic and non-autistic people, and scrutinised against established machine learning algorithms.

The outcomes of this research hold the potential to significantly advance the field of ASD detection by providing more accurate, reliable, and efficient methods. Early and accurate identification of ASD can lead to timely interventions, enabling better treatment outcomes and improved quality of life for children with ASD. Additionally, the proposed approach can be extended to other neurological and developmental disorders that may manifest in distinct facial features.

The structure of the paper is as follows: Section II explores the extant literature on the subject of the paper. Section III discusses the quantum transfer learning model and facial image recognition. Section IV elaborates on the attention mechanisms and transformers for facial image recognition. Section V presents the methodology employed in this research, including the dataset used, experimental setup, and evaluation metrics. Section VI presents the results and discussion of the proposed approach, comparing it with traditional ma- chine learning approaches. Section VII concludes the paper by summarizing. Finally, the acknowledgments section expresses gratitude to individuals or organizations that have contributed to the research.

1. RELATED WORK

Several research have been undertaken to investigate the association between facial expressions and neurological illnesses such as Alzheimer’s [17], neurodegenerative disorders [18], and frontotemporal dementia [19]. In recent years, facial expression analysis in ASD patients has gained popularity. Several strategies have been offered to assist ASD youngsters with processing and performing different facial expressions. Yolcu et al. [20] used a CNN to recognize major ASD-related face characteristics. A second CNN model was employed to recognize the appropriate facial expressions. The greatest accuracy of this model was 94.44 percent. The model was developed further by the same research team by customizing it to recognize characteristics in the lips, eyes, and brows of children with neurological diseases. These models were used to create iconized pictures to help in differential diagnosis.

Valles et al. [21] employed CNN models to analyze human facial expressions using photographs from Kaggle’s (FER2013) 2013 dataset, which was adjusted to incorporate

images of ASD children captured in various lighting situations. Other research [22], [23] employed deep learning and Raspberry Pi3 models to analyze the emotions of ASD children receiving robot-assisted treatment (RAT).

Guha et al. [24] used time-series modeling and statistical analysis to examine six universal emotions in children with and without ASD. Their studies demonstrated that autistic children’s facial expressions are less complex, particularly around the eyes. Grossard et al. [25] created an instructive multimodal emotional imitation game (JEMImE) using a 3- dimensional virtual environment that allows ASD youngsters to express grief, rage, and happiness. To encourage youngsters to play [26], the game incorporated a range of visual and incentive cues.

Smitha et al. [27] examined the effectiveness of parallel and serial PCA feature extraction techniques for a portable motion detector for children with ASD. The PCA technique [28] was implemented in hardware to facilitate the extraction of essential motion characteristics, and it obtained a maximum accuracy of 82.3 percent for a word length of 8 bits. Pramerdorfer et al. [29] looked at using several CNN models to enhance the detection of image-based facial expressions linked with ASD. When compared to other models, the adoption of basic CNN architectures enhanced the detection performance of the program. An ensemble of contemporary deep CNN, on the other hand, attained an accuracy of 75.2%, exceeding previously created models substantially. Another advantage of this model is that it does not require the development of any supplementary face registration or training data.

Ibala et al. [30] used transfer learning and ensemble techniques to identify seven important human emotions in face photographs of persons with and without autism spectrum disorder: sorrow, fear, disgust, neutrality, surprise, happiness, and rage. The greatest accuracy for transfer learning and ensemble approaches was 78.33% and 67.22%, respectively.

1. **FACIAL IMAGE RECOGNITION AND TRANSFORMERS IN ASD DETECTION**

Facial image recognition is crucial for automated Autism Spectrum Disorder (ASD) detection. Attention mechanisms and transformer-based models [52] , known for capturing global dependencies and intricate patterns, are gaining traction in computer vision. This section explores their application in ASD detection via facial image recognition, emphasizing their potential to enhance accuracy and robustness.

While convolutional neural networks (CNNs) traditionally dominated this field [31], [32], [41]–[43], inspired by Natural Language Processing (NLP) success, researchers have fused CNN-like architectures with self-attention [44], [45], and some have even replaced convolutions entirely [46], [47]. Challenges exist in efficiently scaling these models on modern hardware due to specialized attention patterns. Classic ResNet-like architectures remain the state-of-the-art [38], [48]–[50].

This study harnesses Transformers, known for their impact in NLP, to analyze facial images. It partitions images into patches and uses linear embeddings as tokens for supervised image classification. The fusion of attention mechanisms and Transformers holds promise for ASD detection, analyzing facial features, identifying ASD-related patterns, and modeling interactions between facial regions, thereby enhancing discriminative power.

Incorporating attention mechanisms and Transformers into facial image recognition has potential to significantly enhance ASD detection accuracy and robustness, aiding clinicians and researchers in automating screening and early intervention [31], [32], [41]–[43].

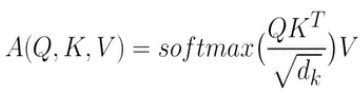
1. ***Transformers for Facial Image Recognition***

Transformers, known for modeling long-range dependencies and capturing context, offer advantages in facial image recognition compared to traditional CNNs.

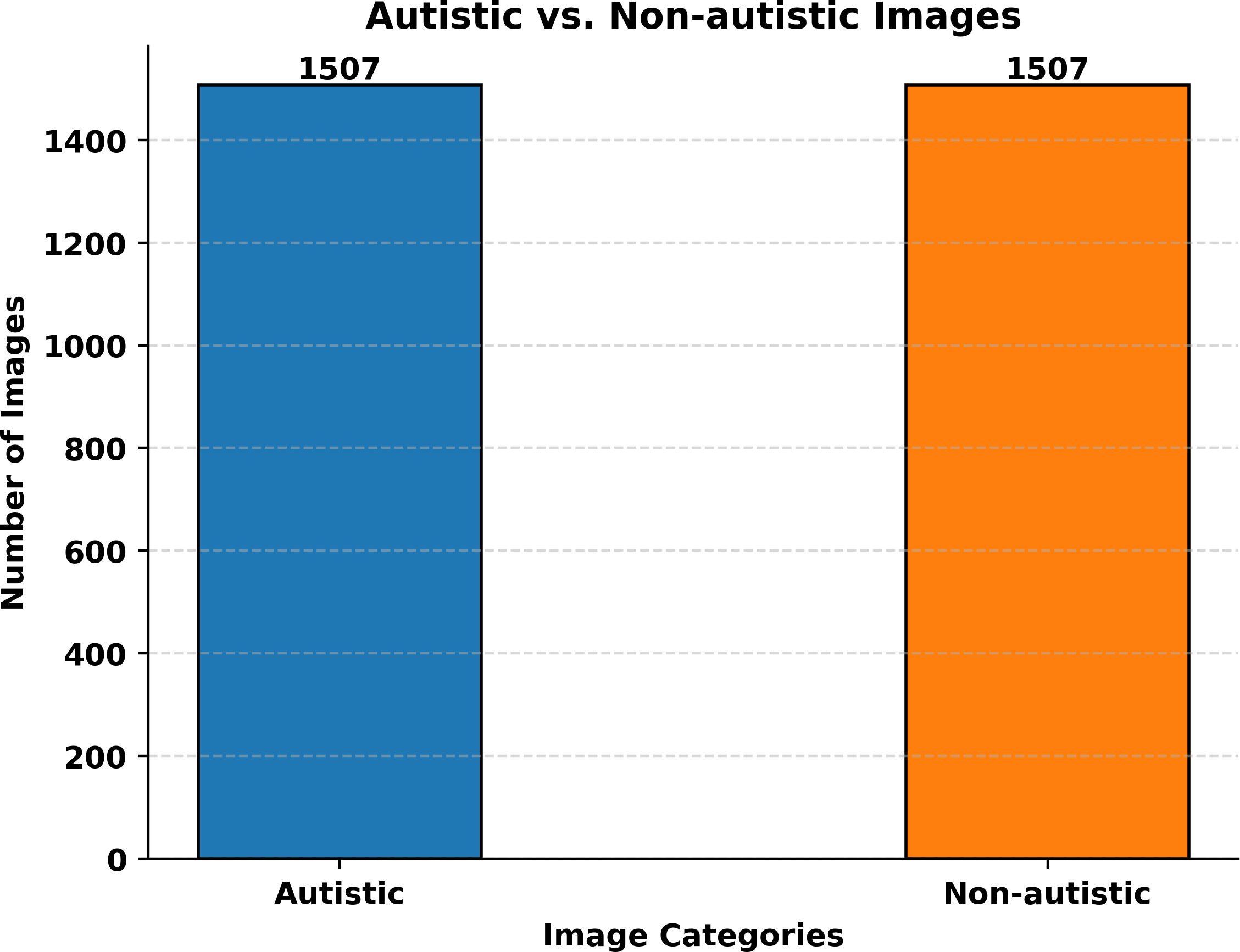
A transformer comprises an encoder and decoder, each with multiple layers. The encoder, vital for visual feature extraction, employs self-attention, feed-forward neural networks, and multi-head attention.

Self-attention allows focusing on pertinent visual informa- To enhance model robustness and generalizability, the dataset tion by comparing query, key, and value vectors:

encompasses diverse age groups, genders, and nationalities.



Attention(𝑄, 𝐾, 𝑉) = (10)



This attention-weighted data undergoes further processing in fully connected layers and non-linear activations, capturing intricate facial patterns and interactions.Multi-head attention enhances the model’s ability to capture diverse relationships by performing parallel self-attention with learned linear projections of queries, keys, and values. The final output combines and linearly transforms the attention-weighted representations from each head.

1. ***Attention in Facial Image Recognition with Transformers***

Transformers utilize attention to focus on relevant facial regions and model their connections, enhancing precision in facial image recognition. Self-attention emphasizes informative areas, capturing facial expressions and landmarks effectively. Multi-head attention further improves discernment by considering diverse dependencies and multiple facets simultaneously. This leads to better recognition of facial expressions and characteristics.

Transformers have shown promise in various computer vision tasks, including face recognition, emotion analysis, and facial attributes. Leveraging attention mechanisms, researchers can develop more accurate and robust facial image analysis models. The next section outlines the research methodology, encompassing the dataset, experimental setup, and evaluation metrics.

1. **PROPOSED METHODOLOGY**

In this section, the methodology employed by the authors for accurate ASD detection through facial image analysis is presented. The first subsection discusses the dataset used for training and evaluation, followed by the data preprocessing steps. Subsequently, the implementation of quantum transfer learning using ResNet-18, ResNet-152, VGG19, and Vision Transformer models is presented.

1. *Dataset Overview*

The dataset used in this study combines author-created data with a publicly accessible Kaggle dataset [32]. It includes 3,014 photos, equally split between autistic and non-autistic children, as seen in Figure 4. Autistic children’s photos were sourced from relevant autism-related sites, while non-autistic children’s photos were randomly collected from the internet.

Fig. 1: Visualization of representative facial images from the dataset.

Figure 1 provides an overview of the number of examples used to evaluate the proposed system, highlighting the distribution of autistic and nonautistic samples for training and testing purposes.

* 1. *Data Preprocessing:* Data preprocessing is crucial for accurate ASD detection via facial image analysis. Steps included normalization for consistent image quality, precise facial region isolation through cropping and alignment, augmentation for robustness, and class balancing via oversampling and undersampling. These measures ensured dataset quality, consistency, and balance, underpinning accurate ASD detection.

1. *Attention-based Quantum Transfer Learning*

The authors present a novel approach that extends transfer learning to hybrid classical-quantum neural networks, enhancing CNN models for image classification. This method follows the general structure of transfer learning but introduces a quantum circuit for classification, a departure from traditional approaches. Figure 2 illustrates the quantum transfer learning progression, with the next section providing a detailed discussion on transfer learning.

* 1. *Selection of Pre-trained Networks:* The selection of appropriate pre-trained networks is a critical step in the attention-based quantum transfer learning framework for ASD detection. These pre-trained networks serve as feature extractors, enabling the model to capture meaningful representations from input images. By leveraging the learned features, the subsequent attention mechanisms and quantum transfer learning algorithms can improve the accuracy of ASD detection.

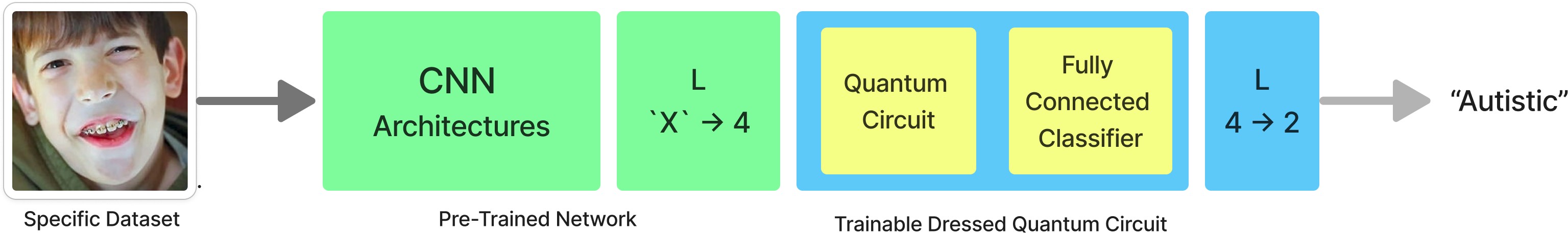


Fig. 2: Flowchart illustrating the quantum transfer learning process.

The ’X’ represents the number of layers in the final convolutional layer of the CNN architecture (e.g., 512, 2048, 1280, etc.).

In this subsection, the selection of three pre-trained networks is discussed: ResNet50, ResNet152, and VGG19. These networks have demonstrated excellent performance in various computer vision tasks and provide valuable feature extraction capabilities.

**ResNet50**: A 18-layer deep CNN by He et al. [1] with residual connections. Pre-trained on ImageNet, it captures vital visual features for ASD detection. Combined with quantum transfer learning, it enhances feature extraction, focusing on relevant facial regions, and improves ASD detection. This fusion boosts model robustness across computer vision tasks, including medical image classification [3].

**ResNet152**: ResNet152, with its 152-layer deep architecture, excels in capturing complex features, makingQuantum Amplitude Encoding: Directly encodes feature values.

* Quantum Phase Encoding: Uses phase modulation for complex relationships.
* Quantum Basis Encoding: Maps features to basis states.
* Quantum Histogram Encoding: Represents features using probability distributions.

For ASD detection, we choose quantum amplitude encoding due to its rich information capture in facial images. The feature map operator is:

it ideal for ASD detection. Pre-trained on ImageNet, it performs exceptionally well in various computer vision tasks.

When integrated into quantum transfer learning, ResNet152 enhances feature extraction, focusing on relevant facial

Quantum kernel method computes inner products between

classical feature vectors with:

regions and enabling comprehensive ASD image analysis, thus improving detection performance.

**VGG19**: VGG19, a deep CNN by Simonyan and Zisserman [1], excels in ASD detection and computer vision. Its 19 layers and uniform 3x3 filters effectively capture features. Pre-trained on ImageNet, VGG19 extracts facial expressions and structures relevant to ASD. Integrated into quantum transfer learning, it improves detection accuracy by focusing on informative regions, emphasizing intricate patterns specific to ASD. Widely adopted and top-performing in image classification challenges [2], VGG19 has a lasting impact on computer vision.

* 1. *Quantum Feature Map:* The quantum feature map con- nects classical input features to quantum computation, enhanc- ing ASD detection. It transforms input features **x** to a quantum state |𝜙(**x**)⟩:

|𝜙(**x**)⟩ = Φ(**x**) |0⟩

Encoding techniques include:

where 𝐾 **x***𝑖*, **x** *𝑗* is the classical kernel function measuring similarity.

( )

In ASD detection, the quantum feature map encodes vital facial information into a quantum state, emphasizing discriminative elements like eye gaze, facial symmetry, and landmarks. It assigns larger amplitudes to indicative ASD features and employs classical kernel functions (e.g., RBF or polynomial kernels) to measure image similarities, effectively capturing ASD-related patterns.

* 1. *Quantum Variational Circuit:* This section delves into the design and structure of the quantum variational circuit used in the proposed methodology. It aims to optimize the ASD detection model’s performance and has been trained on various platforms, including IBM’s quantum computers, QASM simulators, and local machines.

The quantum variational circuit utilizes qubits and quantum gates to explore and optimize quantum states. Qubits are quantum information’s fundamental units, akin to classical bits in classical computing. Quantum gates manipulate qubits for specific computations, allowing complex quantum state creation and manipulation to solve computational proiblems. These gates perform mathematical operations on quantum states for encoding and manipulation.

**Hadamard Gate (H):** The Hadamard gate prepares the qubits in a superposition of states. Mathematically, it is represented as:

Applying the Hadamard gate to each qubit initialized them to

an equal superposition of |0⟩ and |1⟩ states.

**Pauli-Y Rotation Gate (RY)**: The RY gate introduces a rotation angle 𝜃 to each qubit. It is represented as:

By adjusting the rotation angles, the circuit can learn to encode and manipulate the quantum states, allowing it to capture the relevant features necessary for ASD detection.

**Controlled-NOT Gate (CX)**: the CX gate entangles two qubits, here the first qubit acts as the control and the second qubit as the target. When the control qubit is in the state |1⟩, the CX gate flips the state of the target qubit.

**Measurement**: The measurement operation extracts classical information from the quantum system, yielding the outcomes of the qubits.

These mathematical equations represent the fundamental operations performed by the quantum variational circuit.

**Circuit representation**: The mathematical equations for each qubit’s evolution until the first layer of gates are represented as:

For qubit 0 (q0):

The mathematical equation for qubit 0 is:

Adding Ry and CNOT gate:

For qubit 1 (q1):

The mathematical equation for qubit 1 is:

Adding Ry and CNOT gate:

For qubit 2 (q2):

The mathematical equation for qubit 2 is:

Adding and CNOT gate:

For qubit 3 (q3):

The mathematical equation for qubit 3 is:

Adding and CNOT gate :

In the above equations, M represents the measurement operation applied to each qubit after the layer of gates, The symbol and denote the tensor product and identity matrix, respectively , applied to each qubit’s state transformation.

Considering the Hadamard gate applied to all four qubits. The mathematical equation for the Hadamard gate on all qubits is:

Finally, to represent the entire circuit’s mathematical equation, combining all the individual qubit equations up to the first layer of gates, considering the order and combination of operations:

Figure 3 shows the quantum circuit with 6 layers. The equations provided demonstrate the evolution of the circuit up to the first layer, while the remaining layers are added recursively.

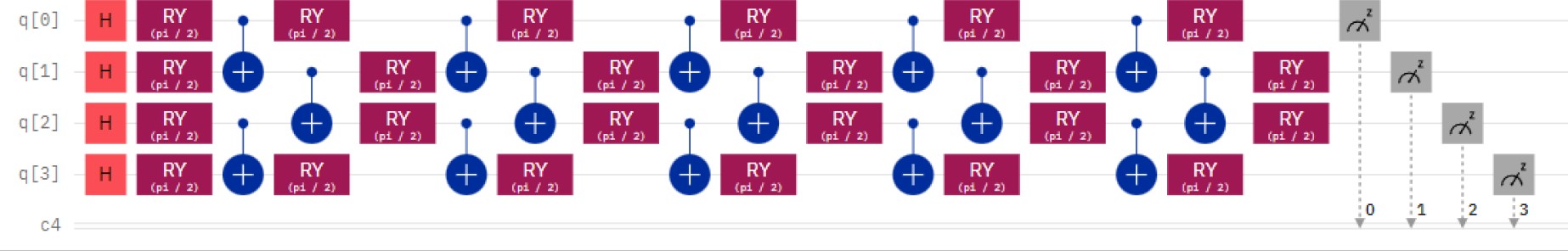


Fig. 3: Quantum Circuit with 6 Layers

Furthermore, the authors discuss the algorithm for defining the quantum layers in the circuit and formulate the quantum net as stand-alone state-of-the-art fully connected layers.

**Algorithm: Defining Quantum Layers in Quantum Cir- cuit** Algorithm 1 defines the quantum layers based on the discussed mathematical equation.

**Algorithm 1** Defining Quantum Layers in Quantum Circuit

Require: The number of qubits, 𝑛𝑞𝑢𝑏𝑖𝑡𝑠

1: function H\_Layer(𝑛𝑞𝑢𝑏𝑖𝑡𝑠) ⊲ /\* Line 1: Layer of single-qubit Hadamard gates \*/

2: for 𝑖𝑑𝑥 0 to 𝑛𝑞𝑢𝑏𝑖𝑡𝑠 − 1 do ⊲ /\* Line 2: Loop over qubits \*/

3: qml.Hadamard(wires = 𝑖𝑑𝑥)

4: end for

5: end function

Require: The rotation angles, 𝑤

6: function R𝑌 𝐿𝑎𝑦𝑒𝑟 W ⊲ /\* Line 6: Layer of parametrized y-axis rotations \*/

7: for 𝑖𝑑𝑥 0 to length(𝑤) − 1 do ⊲ /\* Line 7: Loop over rotation angles \*/

8: qml.RY(𝑤[𝑖𝑑𝑥], wires = 𝑖𝑑𝑥)

9: end for

10: end function

Require: The number of qubits, 𝑛𝑞𝑢𝑏𝑖𝑡𝑠

11: function Entangling\_Layer(𝑛𝑞𝑢𝑏𝑖𝑡𝑠) ⊲ /\* Line 11: Layer of CNOT gates \*/

12: for 𝑖 0 to 𝑛𝑞𝑢𝑏𝑖𝑡𝑠 − 2 by 2 do ⊲ /\* Line 12: Loop over even indices \*/

13: qml.CNOT(wires = [𝑖, 𝑖 + 1])

14: end for

15: for 𝑖 1 to 𝑛𝑞𝑢𝑏𝑖𝑡𝑠 − 2 by 2 do ⊲ /\* Line 15: Loop over odd indices \*/

16: qml.CNOT(wires = [𝑖, 𝑖 + 1])

17: end for

18: end function

**Algorithm: Defining Quantum Net** The quantum net defining algorithm as defined in algorithm 2, describes the formulation of 6 layers variational circuit. The authors further discuss designing dressed quantum net using variational circuits.

* 1. *Dressed Quantum Net for Image Classification:* The dressed quantum net enhances ASD classification accuracy by amalgamating classical CNNs with quantum circuits.

**Algorithm 2** Defining Quantum Net

1: function QUANTUMNET(𝑞𝐼𝑛𝑝𝑢𝑡𝐹𝑒𝑎𝑡𝑢𝑟𝑒𝑠, 𝑞𝑊 𝑒𝑖𝑔ℎ𝑡𝑠𝐹𝑙𝑎𝑡) ⊲ /\* The variational quantum circuit \*/

2: Reshape weights

3: 𝑞𝑊𝑒𝑖𝑔ℎ𝑡𝑠 ← 𝑟𝑒𝑆ℎ𝑎𝑝𝑒

4: Start from state |0⟩, unbiased w.r.t. |0⟩ and |1⟩

5: H\_Layer(𝑛𝑞 𝑞𝑢𝑏𝑖𝑡𝑠)

6: Embed features in the quantum node

7: R𝑌 𝐿𝑎𝑦𝑒𝑟 (QINPUTFEATURES)

8: Sequence of trainable variational layers

9: for 𝑘 0 to 𝑞𝐷𝑒𝑝tℎ 1 do

10: Entangling\_Layer(𝑛𝑞 𝑞𝑢𝑏𝑖𝑡𝑠)

11: R𝑌 𝐿𝑎𝑦𝑒𝑟 (QWEIGHTS[K])

12: end for

13: Expectation values in the 𝑍 basis

14: 𝐸𝑥 𝑝𝑒𝑐𝑡𝑎𝑡𝑖𝑜𝑛 ← [EXPVAL(𝑃𝑎𝑢𝑙𝑖𝑍 (𝑝𝑜𝑠𝑖𝑡𝑖𝑜𝑛)) for 𝑝𝑜𝑠𝑖𝑡𝑖𝑜𝑛(𝑛𝑞 𝑞𝑢𝑏𝑖𝑡𝑠)]

15: return TUPLE(𝐸𝑥 𝑝𝑒𝑐𝑡𝑎𝑡𝑖𝑜𝑛)

16: end function

This integration significantly improves the representation and analysis of image features, harnessing quantum properties such as superposition and entanglement. The quantum layer operates within a hybrid architecture, where classical CNN layers extract lower-level features from the images, and the quantum layer further refines these features using quantum operations. This fusion of classical and quantum approaches demonstrates its effectiveness in improving ASD classification results.

In the context of ASD classification, the dressed quantum net exploits quantum correlations to capture subtle patterns and dependencies in image features. This advanced feature extraction capability enhances the discrimination between ASD and non-ASD samples, showing promise for improved diagnostic accuracy.

To adapt ResNet50, ResNet152, and VGG19, several modifications are applied. These include removing the fully connected layer, adding the quantum layer after convolutional layers, connecting classical CNN outputs to the quantum layer, and freezing classical CNN layer weights. These adjustments optimize the model for the integration of quantum and classical components, resulting in enhanced image classification performance.

**Algorithm: Dressed Quantum Net** Algorithm 3 constructs the fully connected layer by utilizing a quantum circuit. It specifies the information flow from the classical CNN to the quantum fully connected network. On the other hand, Algorithm 4 replaces the classical fully connected layer with a quantum layer. It defines the gradient requirements for the network and determines the number of outputs from the last convolutional layer.

**Algorithm 3** Dressed Quantum Net

1: function DRESSEDQUANTUMNET

2: Initialize pre-processing layer as a linear layer with 512 input neurons and 𝑛𝑞𝑢𝑏𝑖𝑡𝑠 output neurons

3: Initialize quantum parameters as a tensor with 𝑞𝑑𝑒𝑝𝑡ℎ 𝑛𝑞𝑢𝑏𝑖𝑡𝑠 dimensions

4: Initialize post-processing layer as a linear layer with 𝑛𝑞𝑢𝑏𝑖𝑡𝑠 input neurons and 2 output neurons

5: end function

6: function FORWARD(𝑖𝑛𝑝𝑢𝑡\_𝑓𝑒𝑎𝑡𝑢𝑟\_𝑒𝑠)

7: Pass 𝑖𝑛𝑝𝑢𝑡\_𝑓𝑒𝑎𝑡𝑢𝑟\_𝑒𝑠 to pre-processing layer

8: Compute the activation of the pre-processing layer

**Algorithm 4** Model Architecture

1: Input: Chosen-CNN-Network with pretrained weights

2: Output: Model output

3: procedure MAIN

4: Initialize model as 'Chosen-CNN-Network' with pretrained weights

5: for each parameter in the model do

6: Set parameter's requiresGrad to False or True

7: required by the chosen network

8: end for

9: Set the fully connected layer of the model to be a Dressed Quantum Net

10: Move the model to the specified device

11: Obtain input features

12: Get the output by passing the input features through the model

13: return the output

14: end procedure

* 1. *Training of the Network:* The classical-quantum network is trained using a blend of classical and quantum techniques. Key hyperparameters include a 4-qubit quantum layer, a learning rate of 0.0004, a batch size of 4, and a 3-epoch training duration. The quantum circuit depth is set to 6, and a learning rate scheduler reduces the rate by 0.1 every 10 epochs. Quantum weight initialization starts with an initial spread of 0.01. These parameters govern the network’s learning speed, computational complexity, and convergence. This hybrid approach capitalizes on quantum computing’s potential while maintaining compatibility with classical machine learning methods, offering promising solutions for complex problem domains.
  2. *6) Architecture of Vision Transformer:* The Vision Transformer (ViT) revolutionizes computer vision by adopting a transformer-based approach for image analysis. Unlike traditional CNNs, ViT uses self-attention mechanisms to capture intricate image relationships.

ViT’s architecture comprises transformer encoder layers, each with multi-head self-attention and a feed-forward neural network. Self-attention captures long-range dependencies and integrates global context, while the neural network refines features. See Figure 4 for a visual representation.

Algorithm 5 presents an overview of the architecture of the Vision Transformer.

**Algorithm 5** Vision Transformer Architecture

1: **Input:** Image 𝑋, number of transformer layers 𝐿, number of attention heads 𝐻, hidden dimension 𝐷, output dimension 𝑂

2: **Output:** Extracted features 𝑍

3: Initialize input embeddings 𝐸0 with positional encodings

4: **for** 𝑙 = 1 to 𝐿 **do**

5: Apply multi-head self-attention to 𝐸*𝑙*−1 with 𝐻 atten- tion heads

6: Apply feed-forward neural network to the attended features

7: Add skip connections and layer normalization to the output

8: Update 𝐸*𝑙* with the transformed features

9: **end for**

10: Apply global average pooling to 𝐸*𝐿*

11: Linearly project pooled features to obtain 𝑍

12: **return** 𝑍

The algorithm initializes input embeddings with positional encodings for spatial information. Transformer layers apply multi-head self-attention and feed-forward operations iteratively. Skip connections, layer normalization, and global average pooling yield extracted features Z.

The Vision Transformer excels in capturing global and local image relationships using self-attention. It performs exceptionally in various computer vision tasks like image classification, object detection, and semantic segmentation. Its strength in capturing long-range dependencies makes it ideal for tasks requiring deep contextual understanding.

Subsequent sections delve into self-attention and Vision Transformer components, providing insights into its workings and effectiveness in accurate ASD detection through facial image analysis

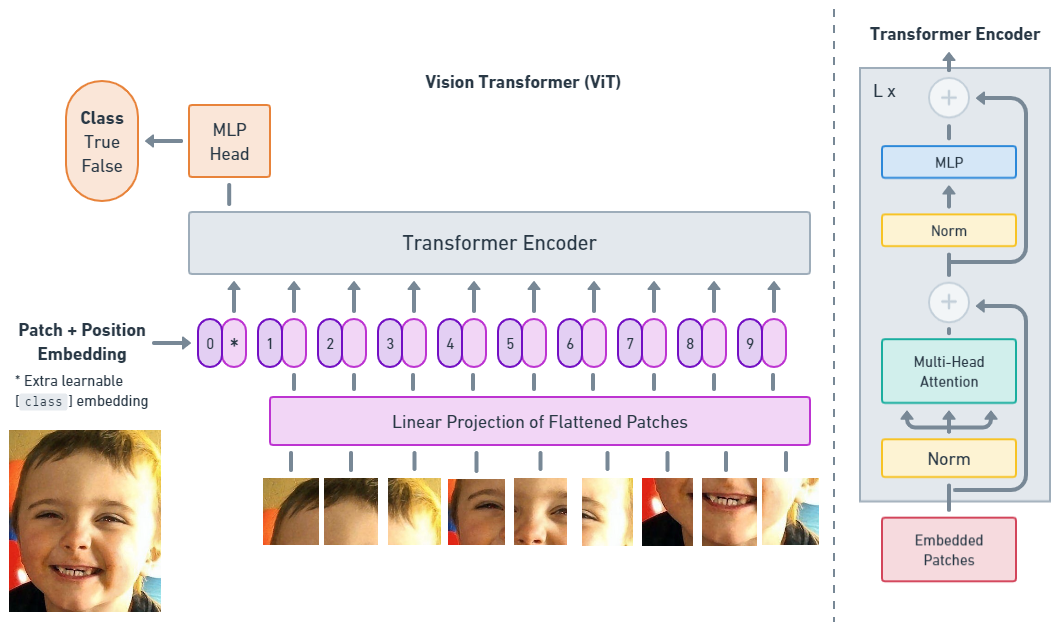


Fig. 4: Model overview.

The image is split into fixed-size patches, linearly embedded, and supplemented with position embeddings. The resulting sequence of vectors is then fed into a standard Transformer encoder. For classification purposes, an additional learnable "classification token" is added to the sequence. The illustration of the Transformer encoder draws Inspiration from [51]

* 1. *Self-Attention Mechanism:* The self-attention mechanism is crucial in the Vision Transformer’s ability to capture complex relationships and dependencies in the autism image dataset for ASD detection.

The input features (**X**) undergo linear transformations to form queries (**Q**), keys (**K**), and values (**V**):

**Q** = **XW***𝑄*, **K** = **XW***𝐾* , **V** = **XW***𝑉*

Attention scores (**A**) are computed through dot products and scaling:

Attended features (Y) result from multiplying attention

scores with the value tensor and applying another linear

transformation:

This self-attention mechanism enhances the model’s ability to extract meaningful facial image features for accurate ASD detection in children. Algorithm 6 provides a summary of this mechanism tailored to ASD detection using facial images.

**Algorithm 6** Classification Head

1: **Input:** Transformed features 𝑍 from the feed-forward network

2: **Output:** Probability distribution over classes 𝑃

3: Linearly project the transformed features 𝑍 to obtain class logits 𝐿

4: Apply softmax activation function to obtain the probability

distribution over classes 𝑃

5: **return** 𝑃

* 1. *Feed-Forward Network:* The Vision Transformer’s feed- forward network extracts meaningful features from attended representations. It comprises two linear layers with non-linear activation (e.g., GELU or ReLU) to capture complex patterns. This process linearly projects attended features, applies activation, and enhances discriminative power in facial image analysis for ASD detection.
  2. *Classification Head:* The classification head is a vital component in the Vision Transformer architecture for the final classification task, distinguishing between autistic and non-autistic children based on extracted features. It operates on transformed features, applying a linear transformation followed by softmax activation to generate probability scores for each class, indicating the image’s category likelihood.

In the classification head, the transformed features Z are linearly projected to obtain class logits L. this projection maps the features to a space where the distances between different classes are better defined, facilitating the classification task. The softmax activation function is then applied to the logits, normalizing them into a probability distribution over the classes. This distribution represents the model’s confidence or uncertainty regarding the classification of the facial image.

The classification head is vital in finalizing the prediction of the Vision Transformer model. It converts the transformed features into a probability distribution, providing a more interpretable output. This enables the model to make informed decisions about classifying autistic and non-autistic children. The probabilities obtained from the classification head help determine the predicted class label and associated confidence level.

In the subsequent sections, the authors will discuss the implementation details of the classification head within the Vision Transformer architecture, including the choice of loss function and training strategies. These details are essential for optimizing the model’s performance and achieving accurate ASD detection using facial image analysis.

* 1. *Training and Hyperparameter Tuning:* Training the Vision Transformer model for ASD detection involves optimizing hyperparameters. The dataset is split into training, validation, and testing sets, ensuring class balance. The number of training epochs is tuned for convergence and overfitting avoidance, and monitored using the validation set. Batch size is adjusted for computational efficiency and stability. Learning rate, another vital hyperparameter, is systematically selected or scheduled for effective training, considering convergence and stability. Regularization techniques, like dropout and weight decay, are applied and fine-tuned to mitigate overfitting while controlling model complexity.

1. **RESULT AND DISCUSSION**

In this section, the authors present the results and discuss the findings of the experiments using quantum transfer learning and vision transformer models for image recognition.

The authors analyze the performance of different pre-trained net- works, including ResNet152, ResNet50, VGG16, and VGG19, in combination with quantum transfer learning. Additionally, the effectiveness of the vision transformer model for image classification is evaluated. The section is organized as follows:

1. *Quantum Transfer Learning Results*

In this subsection, the authors discuss the results of quantum transfer learning using various pre-trained networks.

They present training and validation accuracy and loss graphs for each model. Performance is assessed using metrics like f1-score, precision, recall, and sensitivity.

Figure 6 shows combined model accuracy and loss graphs with eight subplots. The left side represents accuracy, while the right side shows loss. Each subplot tracks the model’s training and validation performance across epochs. The x-axis denotes epochs, and the y-axis indicates accuracy or loss. These subplots offer an overview of training progress, helping assess accuracy and loss fluctuations. They provide insights into model fit and potential adjustments.

Based on the observations of the aforementioned graphs, it can be concluded that among the models evaluated, ResNet50 exhibits the highest accuracy.

Table I presents the individual results of the pre-trained networks, showcasing their performance using various evaluation metrics. The metrics employed include the F1 score, sensitivity, specificity, fall-out, and miss rate. These metrics provide a comprehensive assessment of the networks’ classification capabilities and offer insights into their strengths and weaknesses.

In evaluating various pre-trained networks, ResNet152 achieved the highest accuracy at 0.7640 on the validation dataset, demonstrating strong learning and generalization abilities. ResNet50 followed closely with an accuracy of 0.8661, indicating robust feature capture. VGG19, while slightly lower in accuracy at 0.7002, excelled in top-5 accuracy, showcasing its proficiency in recognizing multiple classes.

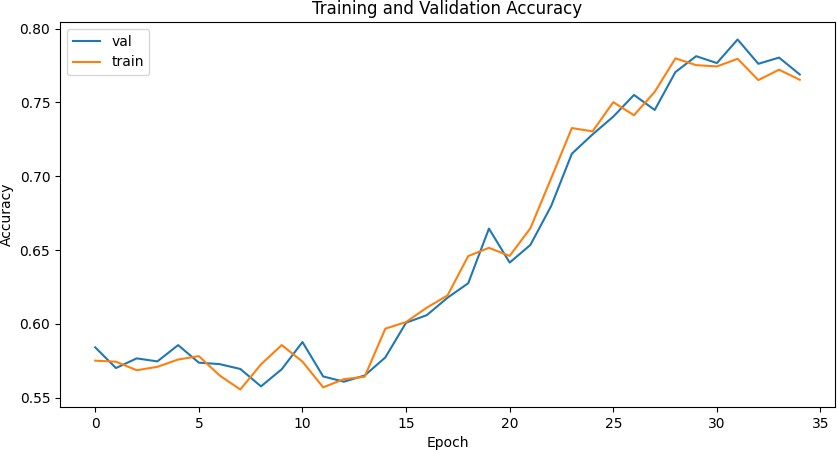
In contrast, VGG16 had the lowest accuracy at 0.5478 and struggled with top-5 accuracy, sensitivity, specificity, fall-out, and miss rate, suggesting limitations in capturing dataset intricacies and complexities.

Further authors discuss about the results of the vision transformer.

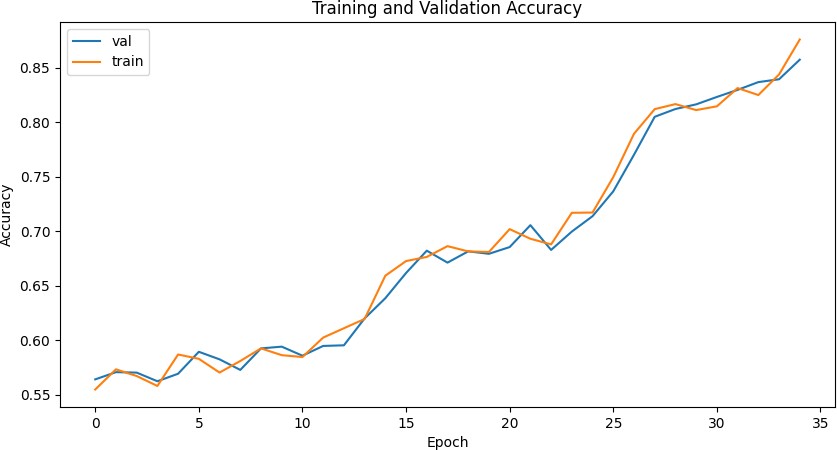
1. *Vision Transformer Results*

In this subsection, the authors focus on the results obtained from the vision transformer model for image recognition. The training and validation accuracy and loss graphs are presented in Figure 7, and the performance of the vision transformer is discussed using the evaluation metrics such as f1-score, precision, recall, and sensitivity.

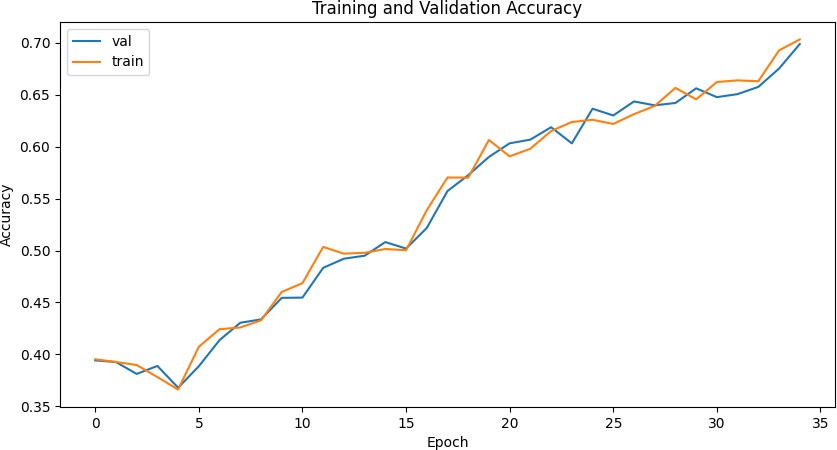
The transformer model exhibits remarkably low loss and achieves a high level of accuracy. Table II presents the experimental results by using the attention-based vision transformer.

ResNet152 Training & Validation Accuracy ResNet152 Training & Validation Loss



ResNet50 Training & Validation Accuracy ResNet50 Training & Validation Loss



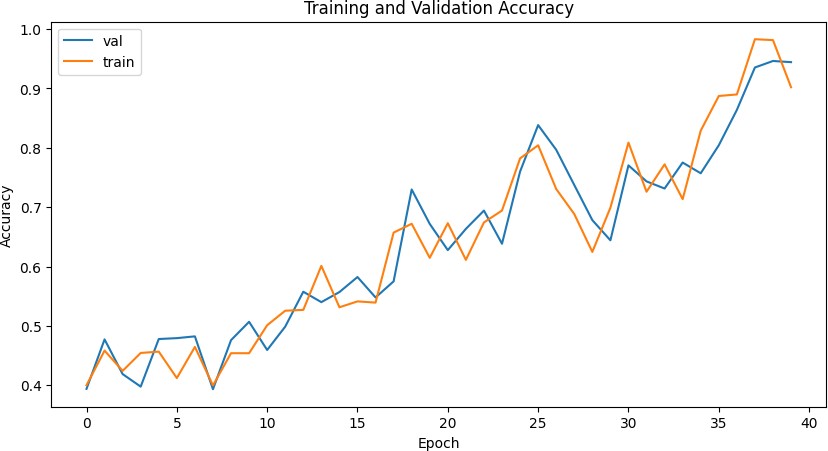
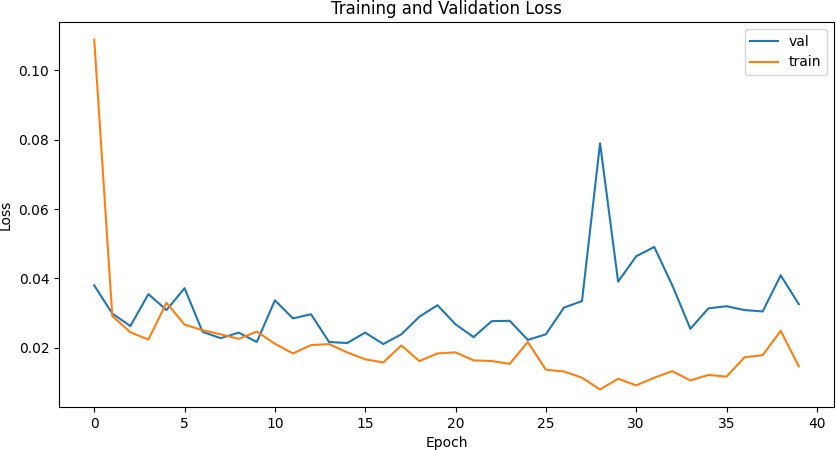
VGG19 Training & Validation Accuracy VGG19 Training & Validation Loss



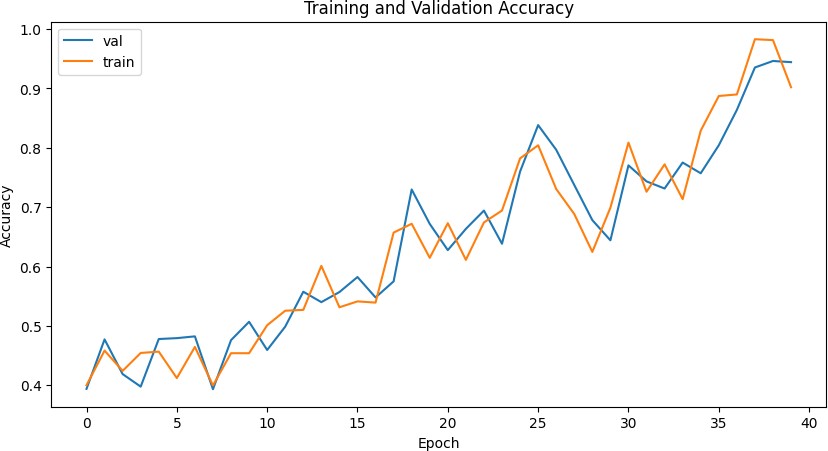
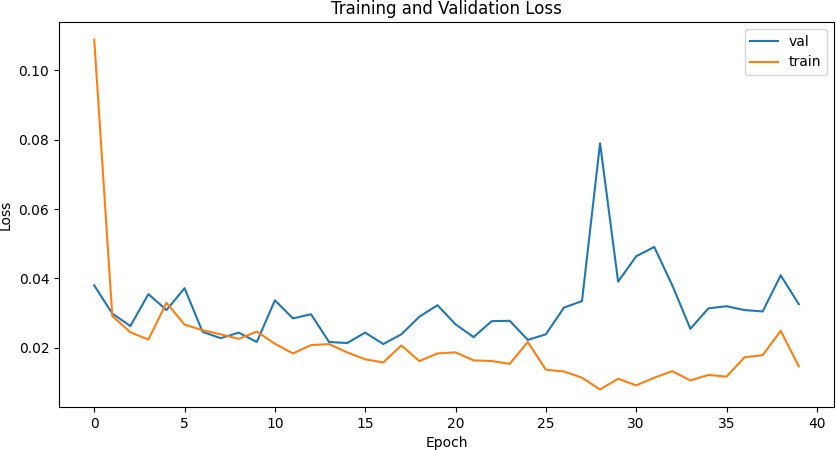
VGG16 Training & Validation Accuracy VGG16 Training & Validation Loss Fig. 5: Training and Validation Accuracy/Loss for Different Models

TABLE I: Experimental results by using the pre-trained models for quantum transfer learning

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Pre-Trained Networks** |  | **Accuracy** |  | **Top-5 Accuracy** | **F-1 Score** | **Sensitivity** | **Specificity** | **Fall-Out** | **Miss Rate** |
|  | **Train** | **Validation** | **Test** |  |  |  |  |  |  |
| ResNet152 | 0.7634 | 0.7640 | 0.7289 | 0.7600 | 0.7631 | 0.7634 | 0.7634 | 0.2366 | 0.2366 |
| ResNet50 | 0.8823 | 0.8661 | 0.8510 | 0.8714 | 0.8833 | 0.8823 | 0.8823 | 0.1177 | 0.1177 |
| VGG19 | 0.7034 | 0.7002 | 0.6933 | 0.6823 | 0.7036 | 0.7034 | 0.7034 | 0.2966 | 0.2966 |
| VGG16 | 0.5534 | 0.5478 | 0.5104 | 0.5162 | 0.5532 | 0.5534 | 0.5534 | 0.4466 | 0.2666 |



Vision Transformer Training & Validation Accuracy Vision Transformer Training & Validation Loss Fig. 6: Comparison of Training and Validation Metrics



Vision Transformer Training & Validation Accuracy Vision Transformer Training & Validation Loss

Fig. 7: Comparison of Training and Validation Metrics

TABLE II: Experimental results by using attention-based vision transformer

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Pre-Trained Networks** | **Accuracy** | **Top-5 Accuracy** | | **F-1 Score** | **Sensitivity** | **Specificity** | **Fall-Out** | **Miss Rate** |
|  | **Train Validation** | **Test** |  |  |  |  |  |  |
| Vision Transformer-Mini | 0.9001 0.8907 | 0.8816 | 0.8671 | 0.8964 | 0.8967 | 0.8967 | 0.0999 | 0.0999 |

1. *Comparative Analysis*

In the comprehensive analysis of the experimental results obtained using attention-based vision transformer and quantum transfer learning, the authors have made several key observations. The attention-based vision transformer, specifically the Vision Transformer-Mini, demonstrates exceptional performance with high accuracy, top-5 accuracy, and F-1 score. The experimental results, as shown in Table II, indicate that the Vision Transformer-Mini achieves a training accuracy of 0.9001, a validation accuracy of 0.8907, and a test accuracy of 0.8816. Furthermore, it achieves a top-5 accuracy of 0.8671, highlighting its ability to provide accurate predictions among the top five predicted classes. The F-1 score of 0.8964 further emphasizes its robust performance. Moreover, the Vision Transformer-Mini exhibits favorable sensitivity, specificity, fall-out, and miss rate values, with each metric close to 0.1. These metrics reflect a well-balanced performance in correctly identifying positive and negative instances and minimizing false positives and false negatives.

It is important to note that the Quantum Transfer Learning model using the ResNet50 architecture also demonstrates promising results. The ResNet50 model showcases performance similar to the Vision Transformer- Mini. Specifically, the ResNet50 model achieves competitive accuracy, top-5 accuracy, and F-1 score values, indicating its effectiveness in image classification tasks. Considering the similarity in performance between the Quantum Transfer Learning model with ResNet50 and the Vision Transformer-Mini, it becomes evident that the ResNet50 model can serve as a viable alternative. The ResNet50 model is particularly noteworthy in the context of quantum transfer learning, showcasing its potential for leveraging quantum techniques to enhance image classification capabilities. The attention-based vision transformer and the Quantum Transfer Learning model with ResNet50 both exhibit strong performance, underscoring their efficacy in image classification tasks. These results offer valuable insights into the effectiveness of various models and approaches, empowering researchers and practitioners to make informed choices when selecting models for their specific applications.

1. **CONCLUSION**

In conclusion, the study introduces an innovative framework that synergizes quantum transfer learning with transformer models to elevate machine learning tasks. This framework seamlessly integrates the Vision Transformer with pre-trained CNN architectures like ResNet50, ResNet152, VGG19, and VGG16 for quantum transfer learning. Extensive experiments validate the framework’s superiority over traditional transfer learning methods, delivering substantial gains in accuracy, convergence speed, and generalization. It adeptly captures inherent patterns from pre-trained CNNs while employing the Vision Transformer to enhance sequential data processing, focusing on pertinent features and long-range dependencies.

Moreover, optimization via Intel’s oneDNN library enhances computational efficiency, accelerating convergence. This amalgamation of quantum transfer learning, the Vision Transformer, and streamlined training processes yields a versatile and efficient framework adaptable across diverse domains and datasets, enhancing learning performance and scalability. This research’s potential to inspire further exploration lies at the intersection of quantum computing, transfer learning, transformer models, and optimization techniques. The framework presents promising avenues to address

complex machine learning challenges, especially in scenarios featuring limited large-scale datasets or sequential data, all while harnessing the exponential computational capabilities of quantum computing. Researchers can also focus on creating specialized datasets optimized for quantum transfer learning while integrating quantum transformations seamlessly into the data augmentation process.

1. REFERENCES
2. M. S. Satu, F. F. Sathi, M. S. Arifen, M. H. Ali, and M. A. Moni, “Early detection of autism by extracting features: a case study in bangladesh,” in *2019 international conference on robotics, electrical and signal processing techniques (ICREST)*. IEEE, 2019, pp. 400–405.
3. Q. Guillon, N. Hadjikhani, S. Baduel, and B. Rogé, “Visual social at- tention in autism spectrum disorder: Insights from eye tracking studies,” *Neuroscience & Biobehavioral Reviews*, vol. 42, pp. 279–297, 2014.
4. M. I. U. Haque and D. Valles, “A facial expression recognition approach using dcnn for autistic children to identify emotions,” in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, 2018, pp. 546–551.
5. O. Rudovic, Y. Utsumi, J. Lee, J. Hernandez, E. C. Ferrer, B. Schuller, and R. W. Picard, “Culturenet: A deep learning approach for engagement intensity estimation from face images of children with autism,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 339–346.
6. G. Yolcu, I. Oztel, S. Kazan, C. Oz, K. Palaniappan, T. E. Lever, and

F. Bunyak, “Facial expression recognition for monitoring neurological disorders based on convolutional neural network,” *Multimedia Tools and Applications*, vol. 78, pp. 31 581–31 603, 2019.

1. T. Akter, M. S. Satu, L. Barua, F. F. Sathi, and M. H. Ali, “Statistical analysis of the activation area of fusiform gyrus of human brain to explore autism,” *Int. J. Comput. Sci. Inf. Secur.(IJCSIS)*, vol. 15, pp. 331–337, 2017.
2. M. S. Satu, M. S. Azad, M. F. Haque, S. K. Imtiaz, T. Akter, L. Barua,

M. Rashid, T. R. Soron, and K. A. Al Mamun, “Prottoy: A smart phone based mobile application to detect autism of children in bangladesh,” in *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*. IEEE, 2019, pp. 1–6.

1. S. Schelinski, K. Borowiak, and K. von Kriegstein, “Temporal voice areas exist in autism spectrum disorder but are dysfunctional for voice identity recognition,” *Social Cognitive and Affective Neuroscience*, vol. 11, no. 11, pp. 1812–1822, 2016.
2. M. D. Hossain, M. A. Kabir, A. Anwar, and M. Z. Islam, “Detecting autism spectrum disorder using machine learning techniques: An exper- imental analysis on toddler, child, adolescent and adult datasets,” *Health Information Science and Systems*, vol. 9, pp. 1–13, 2021.
3. F. I. ADVERTISEMENTFEATURE, “Quantum imaging making person- alised medicine a reality.”
4. P. Jayanthi, B. K. Rai, and I. Muralikrishna, “The potential of quantum computing in healthcare,” in *Technology Road Mapping for Quantum Computing and Engineering*. IGI Global, 2022, pp. 81–101.
5. R. Ur Rasool, H. F. Ahmad, W. Rafique, A. Qayyum, J. Qadir, and

Z. Anwar, “Quantum computing for healthcare: A review,” *Future Internet*, vol. 15, no. 3, p. 94, 2023.

1. T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi,

P. Cistac, T. Rault, R. Louf, M. Funtowicz *et al.*, “Transformers: State- of-the-art natural language processing,” in *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, 2020, pp. 38–45.

1. D. Rothman, *Transformers for Natural Language Processing: Build innovative deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more*. Packt Publishing Ltd, 2021.
2. J. Bi, Z. Zhu, and Q. Meng, “Transformer in computer vision,” in *2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*. IEEE, 2021, pp. 178–188.
3. A. Parvaiz, M. A. Khalid, R. Zafar, H. Ameer, M. Ali, and M. M. Fraz, “Vision transformers in medical computer vision—a contemplative retrospection,” *Engineering Applications of Artificial Intelligence*, vol. 122, p. 106126, 2023.
4. T. M. Ghazal, S. Abbas, S. Munir, M. Khan, M. Ahmad, G. F. Issa, S. B. Zahra, M. A. Khan, and M. K. Hasan, “Alzheimer disease detection empowered with transfer learning.” *Computers, Materials & Continua*, vol. 70, no. 3, 2022.
5. M. D. Sweeney, K. Kisler, A. Montagne, A. W. Toga, and B. V. Zlokovic, “The role of brain vasculature in neurodegenerative disorders,” *Nature neuroscience*, vol. 21, no. 10, pp. 1318–1331, 2018.
6. D. Fernandez-Duque and S. E. Black, “Impaired recognition of negative facial emotions in patients with frontotemporal dementia,” *Neuropsy- chologia*, vol. 43, no. 11, pp. 1673–1687, 2005.
7. G. Yolcu, I. Oztel, S. Kazan, C. Oz, K. Palaniappan, T. E. Lever, and

F. Bunyak, “Facial expression recognition for monitoring neurological disorders based on convolutional neural network,” *Multimedia Tools and Applications*, vol. 78, pp. 31 581–31 603, 2019.

1. M. I. U. Haque and D. Valles, “A facial expression recognition approach using dcnn for autistic children to identify emotions,” in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, 2018, pp. 546–551.
2. A. Di Nuovo, D. Conti, G. Trubia, S. Buono, and S. Di Nuovo, “Deep learning systems for estimating visual attention in robot-assisted therapy of children with autism and intellectual disability,” *Robotics*, vol. 7, no. 2, p. 25, 2018.
3. S. Singh, R. Ramya, V. Sushma, S. Roshini, and R. Pavithra, “Facial recognition using machine learning algorithms on raspberry pi,” in *2019 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT)*. IEEE, 2019, pp. 197–202.
4. T. Guha, Z. Yang, A. Ramakrishna, R. B. Grossman, D. Hedley,

S. Lee, and S. S. Narayanan, “On quantifying facial expression-related atypicality of children with autism spectrum disorder,” in *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2015, pp. 803–807.

1. C. Grossard, S. Hun, A. Dapogny, E. Juillet, F. Hamel, H. Jean-Marie,

J. Bourgeois, H. Pellerin, P. Foulon, S. Serret *et al.*, “Teaching facial expression production in autism: The serious game jemime,” *Creative Education*, vol. 10, no. 11, p. 2347, 2019.

1. A. Dapogny, C. Grossard, S. Hun, S. Serret, J. Bourgeois, H. Jean- Marie, P. Foulon, H. Ding, L. Chen, S. Dubuisson *et al.*, “Jemime: a serious game to teach children with asd how to adequately produce facial expressions,” in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 723–730.
2. K. G. Smitha and A. P. Vinod, “Facial emotion recognition system for autistic children: a feasible study based on fpga implementation,” *Medical & biological engineering & computing*, vol. 53, pp. 1221–1229, 2015.
3. S. X. Wu, H.-T. Wai, L. Li, and A. Scaglione, “A review of distributed algorithms for principal component analysis,” *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1321–1340, 2018.
4. S. Singh and F. Nasoz, “Facial expression recognition with convolutional neural networks,” in *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2020, pp. 0324–0328.
5. R. Kaly-Ibala, J. Mabiala-Babela, and P. Senga, “Encephalopathy after pertussis immunization,” *ARCHIVES DE PEDIATRIE*, vol. 7, no. 2, pp. 216–+, 2000.
6. J. Heaton, “Ian goodfellow, yoshua bengio, and aaron courville: Deep learning: The mit press, 2016, 800 pp, isbn: 0262035618,” *Genetic Programming and Evolvable Machines*, vol. 19, no. 1-2, pp. 305–307, 2018.
7. E. Farhi and H. Neven, “Classification with quantum neural networks on near term processors,” *arXiv preprint arXiv:1802.06002*, 2018.
8. J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, “The theory of variational hybrid quantum-classical algorithms,” *New Journal of Physics*, vol. 18, no. 2, p. 023023, 2016.
9. A. Perdomo-Ortiz, M. Benedetti, J. Realpe-Gómez, and R. Biswas, “Opportunities and challenges for quantum-assisted machine learning in near-term quantum computers,” *Quantum Science and Technology*, vol. 3, no. 3, p. 030502, 2018.
10. J. Tung, E. A. Archie, J. Altmann, and S. C. Alberts, “Cumulative early life adversity predicts longevity in wild baboons,” *Nature communica- tions*, vol. 7, no. 1, p. 11181, 2016.
11. M. Schuld and N. Killoran, “Aprendizaje automático cuántico en espa- cios característicos de hilbert,” *Cartas de revisión física*, vol. 122, p. 040504, 2019.
12. M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, “Circuit-centric quantum classifiers,” *Physical Review A*, vol. 101, no. 3, p. 032308, 2020.
13. N. Killoran, T. R. Bromley, J. M. Arrazola, M. Schuld, N. Quesada, and S. Lloyd, “Continuous-variable quantum neural networks,” *Physical Review Research*, vol. 1, no. 3, p. 033063, 2019.
14. S. S. J. P. Aspuru-Guzik, “A expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms adv,” *Quant. Technol*, vol. 2, no. 12, p. 1900070, 2019.
15. J. Preskill, “Quantum computing in the nisq era and beyond,” *Quantum*, vol. 2, p. 79, 2018.
16. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard,

W. Hubbard, and L. D. Jackel, “Backpropagation applied to handwritten zip code recognition,” *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.

1. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
2. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition. cvpr. 2016,” *arXiv preprint arXiv:1512.03385*, 2016.
3. X. Wang, R. Girshick, A. Gupta, and K. He, “Non-local neural net- works,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7794–7803.
4. N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and

S. Zagoruyko, “End-to-end object detection with transformers,” in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*. Springer, 2020, pp. 213–229.

1. P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, and

J. Shlens, “Stand-alone self-attention in vision models,” *Advances in neural information processing systems*, vol. 32, 2019.

1. H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille, and L.-C. Chen, “Axial-deeplab: Stand-alone axial-attention for panoptic segmentation,” in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV*. Springer, 2020, pp. 108–126.
2. D. Mahajan, R. Girshick, V. Ramanathan, K. He, M. Paluri, Y. Li,

A. Bharambe, and L. Van Der Maaten, “Exploring the limits of weakly supervised pretraining,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 181–196.

1. A. Kolesnikov, L. Beyer, X. Zhai, J. Puigcerver, J. Yung, S. Gelly, and

N. Houlsby, “Big transfer (bit): General visual representation learning,” in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*. Springer, 2020, pp. 491–507.

1. Q. Xie, M.-T. Luong, E. Hovy, and Q. V. Le, “Self-training with noisy student improves imagenet classification,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10 687–10 698.
2. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.
3. S. Karthikeyan, T. Banerjee,Amandeep Sharma and Charvi et.al(14 Aug 2021), " Attention based Discrimination of Mycoplasma Pneumonia," Springer 4th International Conference on Computational Intelleigence and *Data Engineering(ICCIDE-2021), VIT University,AP.pp.29-https://doi.org/10.1007/978-981-16-*