# Diagnosing a Child with Autism using Artificial Intelligence

**Abdulrahman Alharbi[1] , Hadi Alyami, Saleh Alenzi, Saud Alharbi, Zaid bassfar**

*University of Tabuk Department of Information Technology, TU, KSA*

**Abstract**
Children are the foundation and future of this society and understanding their impressions and behaviors is very important and the child's behavioral problems are a burden on the family and society as well as have a bad impact on the development of the child, and the early diagnosis of these problems helps to solve or mitigate them, and in this research project we aim to understand and know the behaviors of children, through artificial intelligence algorithms that helped solve many complex problems in an automated system, By using this technique to read and analyze the behaviors and feelings of the child by reading the features of the child's face, the movement of the child's body, the method of the child's session and nervous emotions, and by analyzing these factors we can predict the feelings and behaviors of children from grief, tension, happiness and anger as well as determine whether this child has the autism spectrum or not. The scarcity of studies and the privacy of data and its scarcity on these behaviors and feelings limited researchers in the process of analysis and training to the model presented in a set of images, videos and audio recordings that can be connected, this model results in understanding the feelings of children and their behaviors and helps doctors and specialists to understand and know these behaviors and feelings.

***Keywords:***
*.Autism"ASD", Children, Machine learning , CNN , RNN*

## 1. Introduction

### 1.1 Definition

Autism Spectrum Disorder (ASD) is a brain development disorder that affects how a person discriminates and interacts with others on a social level, producing social communication and interaction issues. Limited and recurring behavioral patterns are also part of the condition. The term "spectrum" refers to a wide variety of symptoms and severity levels in autism spectrum disorder.

Autism can be also defined as a disorder affecting children in the first three years of life where the disorder includes the inability of the child to establish meaningful social relationships, that he suffers from cognitive disorder and poor motivation and has a defect in the development of cognitive functions and the inability to understand the concepts of intent and spatiality and has a severe inability to use and develop language and that he/she suffers from any class of stereotypical play and poor ability to imagine and resist changes in his environment.

### 1.2 Study background

Complications (problems caused by autism) can be Problems related to social interaction, communication and behavior can lead to Problems at school related to successful learning, Inability to live independently. Social isolation, psychological stress within the family, and Victimization and bullying. While the target group is the children aged 3-17 so that early detection of autism will be useful in finding a cure for the infected person.

### 1.3 Significance of study

Early diagnosis of autism spectrum disorder (ASD) has been linked to considerable gains in intellectual ability, adaptive behavior, and symptom severity in children with ASD. According to studies, the use of autistic children's behavioral clues is a frequent way of ASD diagnosis. ASD is a group of neurodevelopmental disorders marked by difficulty with social communication and interaction, as well as repetitive behaviors and narrow interests. The prevalence of the disease has been rising, and current diagnostic approaches are both time and labor intensive.

Manuscript received June 5, 2022
Manuscript revised June 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.6.22

### 1.4 Scope of the study

The Autism Diagnostic Observation Schedule (ADOS) and the Autism Observation Scale (AOS). for Infants are two testing instruments that use behavioral signals to diagnose ASD (AOS). Furthermore, self-stimulatory behaviors are abnormal behavioral cues that are tested for diagnosis in these instruments. To diagnose autism, physicians must interact with the child over several long sessions to uncover behavioral signs.

In other locations, however, appropriately educated clinicians may be scarce and costly. As a result, utilizing a computer to analyze features of children with autism, such as self-stimulatory behavior, can assist clinicians in making a more accurate diagnosis. Self-stimulatory behaviors like head pounding are categorized as self-injurious because they potentially harm children.

### 2.5 Objectives

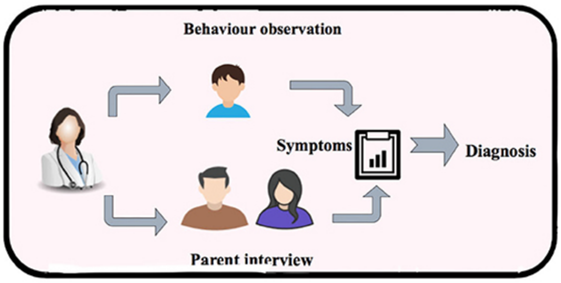
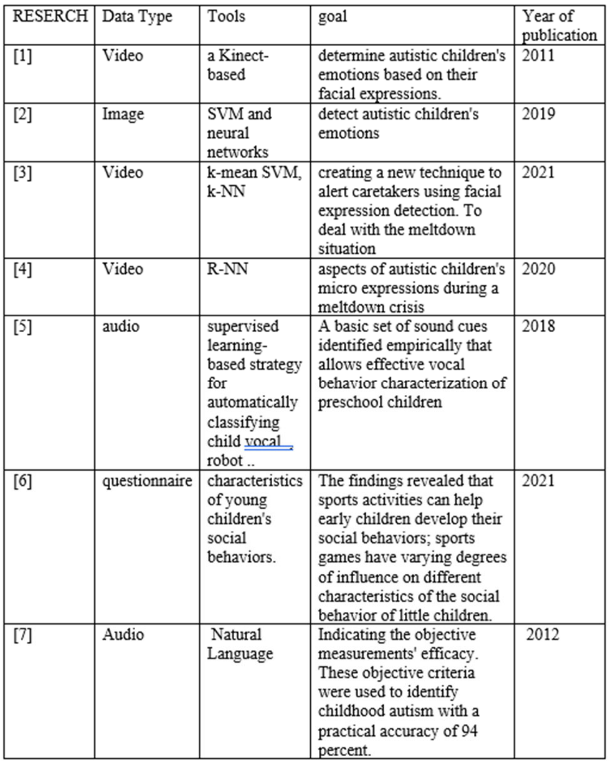


Fig. 1 Autism detection.

It is impractical to study autistic children at all hours of the day due to the random occurrence of self-stimulatory behavior. Doctors and parents of children with autism may benefit from an automatic self-stimulatory behavior analysis system.

We will validate this model to be used in children's clinics to help doctors diagnose as well as in primary schools to help teachers and mentors understand children's feelings, strange behaviors, and sudden emotions at school as shown in figure 1.

## 2. Related Work



| RESERCH | Data Type | Tools | goal | Year of publication |
|---|---|---|---|---|
| [1] | Video | a Kinect-based | determine autistic children's emotions based on their facial expressions. | 2011 |
| [2] | Image | SVM and neural networks | detect autistic children's emotions | 2019 |
| [3] | Video | k-mean SVM, k-NN | creating a new technique to alert caretakers using facial expression detection. To deal with the meltdown situation | 2021 |
| [4] | Video | R-NN | aspects of autistic children's micro expressions during a meltdown crisis | 2020 |
| [5] | audio | supervised learning-based strategy for automatically classifying child vocal robot .. | A basic set of sound cues identified empirically that allows effective vocal behavior characterization of preschool children | 2018 |
| [6] | questionnaire | characteristics of young children's social behaviors. | The findings revealed that sports activities can help early children develop their social behaviors; sports games have varying degrees of influence on different characteristics of the social behavior of little children. | 2021 |
| [7] | Audio | Natural Language | Indicating the objective measurements' efficacy. These objective criteria were used to identify childhood autism with a practical accuracy of 94 percent. | 2012 |

Fig. 2 Comparison between Related Work

Current research on self-stimulatory behavior is mostly separated into two categories: accelerometer-based research and computer vision-based research. Because 2D cameras are less expensive and more accessible, we chose to build a self-stimulatory behavior classification system based on video data.

In a study by [1], tantrums have a negative impact on a child's development. It is critical to supply doctors with authentic and trustworthy data for them to detect problematic behaviors in children in a timely manner. Children's behavior analysis via video has been presented as a solution; nevertheless, existing implementations face three challenges: caregiver attitudes, video collection and analysis, and techniques to proving their efficacy. To address these issues, the authors of a research suggested a prototype system that incorporated medical knowledge, questionnaire-based attitudes analysis, and a Kinect-based behavior analysis algorithm. They designed a questionnaire to interview the doctor and analyses the feedback results to evaluate the method's effectiveness.Facial Emotion Detection is a method of recognizing human emotions by analyzing facial expressions. Autism Spectrum Disorder (ASD) is a severe neurobehavioral condition. Autistic persons have a habit of acting in the same way repeatedly. They aren't ready to engage in social interaction. This syndrome affects people's ability to recognize emotions. The goal of the study was to determine autistic children's emotions based on their facial expressions. A study in [2] focused on four different emotions. Sad, joyful, neutral, and angry are the four emotions. Image processing and machine learning techniques are used to detect autistic children's emotions. The characteristics were retrieved using a local binary pattern from the faces of autistic children. Emotions are classified using machine learning algorithms. neural networks and Support vector machines are two machine learning classifiers that were utilized in the classification process.

Autism spectrum disorder (ASD) is a symptom that affects many people. Autism must be diagnosed as soon as possible to have a good prognosis. The most prevalent approach of diagnosis employs autistic children's behavior clues. Years of clinical training are required for doctors to be able to recognize these behavioral cues (such as self-stimulatory behaviors). Artificial intelligence technology has been able to automatically capture self-stimulatory behaviors because to advances in deep learning algorithms and hardware in recent years. Doctors' work efficacy can be increased with this strategy. However, enough labeled dataset to train a model is currently lacking in the field of self-stimulatory behavior study. A study in [3] used the temporal coherency between each neighboring frame as free supervision and the establishment of a global discriminative margin to extract slow-changing discriminative self-stimulatory behavior. The usefulness of the extracted features has been confirmed by extensive evaluation. To demonstrate the classification of self-stimulation behaviors in an unsupervised manner, the retrieved characteristics are first classified using the k-means approach. The usefulness of features is then assessed using the conditional entropy approach. Second, we use a hybrid approach using both the unsupervised TCDN approach and the supervised learning optimization techniques to achieve good results. These cutting-edge findings demonstrate the efficacy of slow-changing discriminative self-stimulatory behavior traits.

Recognizing human emotion has made a significant contribution to computer vision applications. This effort, despite its importance, emphasizes the safety of autistic persons in meltdown crises by creating a new technique to alert caretakers using facial expression detection. To deal with the meltdown situation, a preventive approach has been implemented. Meltdown symptoms are unquestionably linked to aberrant facial expressions associated with complex emotions. In fact, researchers previously believed that Human Facial Expressions (HFE) could only represent the seven basic emotions. HFE is a complex emotion that might suggest two or more emotions, known as compound or mixed emotions, according to psychologists. Compound Emotion has been the subject of a few investigations (CE). In addition, there are several tough challenges to detect Compound Emotion Recognition (CER). In a study by [4], the researchers experimentally evaluated a set of deep spatiotemporal geometric aspects of autistic children's micro expressions during a meltdown crisis. To do this, researchers compared CER performance with a variety of micro-expression variables to identify the aspects that best distinguish autistic children CE in meltdown crisis from normal condition, as well as the best classifier performance. Researchers used the

Kinect camera to record films of autistic youngsters in normal and meltdown situations. The experimental results revealed that employing Information Gain Feature Selection techniques, with both deep spatial-temporal features that represent geometry, and the Recurrent Neural Network RNN to provide the greatest performance (85.8 percent).

Autism is a neurodevelopmental illness that affects a growing number of youngsters and has serious social and economic ramifications for those who are affected and their families. Incorporating robotic technologies into the diagnosis process could improve its speed and accuracy, paving the path for earlier and more effective treatment. The diagnostic approach necessitates multimodal engagement, with the child's vocal behavior playing a key part. A study in [5] described a supervised learning-based strategy for automatically classifying child vocal behavior that is suited for real-time execution on an autonomous robot with limited computational resources. The study's main contribution was an empirically discovered basic collection of sound cues that may be used to classify pre-school children's vocal behavior that is valuable in autism diagnosis. The classifier was put to the test on a dataset that contained both publicly available audio recordings and recordings from therapeutic and diagnostic sessions.

People's attitudes, behavioral responses, and language, to another or communication activities are referred to as social behaviors. A study in [6] combined a questionnaire and an experimental approach to evaluate the effective ways and means of the development little children's social behavior, using sports activities as the training variables, from two perspectives of little children competency evaluation and issue behavior evaluation. The findings revealed that sports activities can help early children develop their social behaviors; sports games have varying degrees of influence on different characteristics of young children's social behaviors; and different age stages of young children's social behavior development.

Emotional or behavioral difficulties in children, as well as inadequate social development, have been a significant burden on families and society. However, no large-scale studies on emotional and behavioral disorders, as well as social skills, are currently available. A total of 9,295 students aged 6 to 16 years old were enrolled in a study in China [7]. Students were assessed for emotional and behavioral disorders as well as social abilities using the Child Behavior Checklist (CBCL). Then the study looked at the elements that were significant predictors of children's behavioral difficulties and social competencies. Children with behavioral disorders scored considerably worse on social and learning abilities than children without behavioral problems (P=0.05). Gender, developmental delay, recent life experiences, unfavorable relationships, and bad child-rearing approaches were all common contributing variables for behavioral issues and social competence. Non-breastfeeding, age, macrosomia, threatening abortion, hospitalization for physical sickness, physical illness, poor sleep, and non-breastfeeding were all found to be independent risk factors for emotional and behavioral problems in children. Students in Beijing have major social competency, emotional, and behavioral issues. Mental health should be given more emphasis, and effective intervention tools should be made available.

This work [8] tries to solve the problem by delving into the depths of underlying characteristics and revealing hidden behavioral information in audio streams. Using signal processing and pattern recognition methods on daylong audio recordings led to the discovery that most of a child's verbal behavior may be evaluated automatically. The model obtains an overall autism detection accuracy of 94 percent (N=226) by incorporating numerous such variables. This strategy, like many other developing non-invasive and telemonitoring health-care technologies, is seen to have enormous potential in child development research, clinical practice, and parenting.

## 3. Methodology

To distinguish those with ASD from those without ASD, training a classification algorithm with automatically and objectively measured features from many autistic and other generally developed individuals is used. A machine learning technique is particularly well suited to capturing heterogeneous and complicated behavior in real-world social interactions, and it will be crucial in the development of automated and objective categorization systems for ASD.

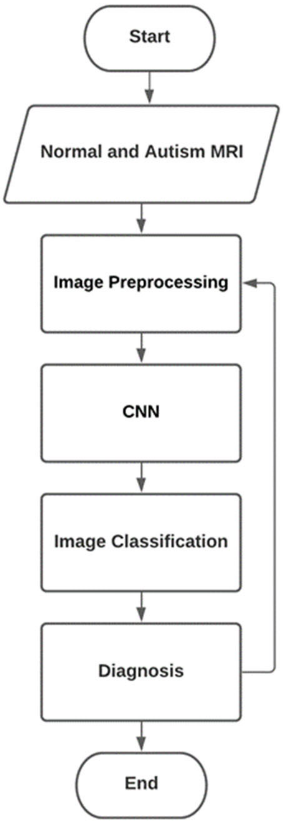


Fig. 3 Research methodology

As a result, our proposed paradigm is a proof-of-concept analysis of data from a social interaction research to determine the feasibility of classifying autistic persons using full-body nonverbal behavior data collected from cameras capturing naturalistic social interactions from both images and videos of children with autism and normal children.

The research methodology illustrated in figure 2 proposed a few steps starting with images of normal and autism children with equal numbers to be balanced. The second step is to preprocess the acquired images to filter, crop, and resize them. The third step is to apply a suitable Convolutional Neural Network for labeled dataset to extract the features and classify the children's images. Classification is performed and validated with suitable accuracy measures to select the best deep learning model. An additional step of utilizing videos capturing the social interactivity of children suffering from autism as well as videos capturing normal children social behavior. That stage will use a combination of CNN and RNN to accurately predict the change of behavior and the discrepancies between normal children and children with autism.

Most of the machine learning (ML) investigations in ASD research have used simple cross-validation (CV) approaches. This raises the chances of selecting an overly optimistic model. As a result, we recommend using a second layer of CV to avoid overfitting by allowing parameter selection and model performance evaluation to be done separately on different data sets. The test fold is totally held out until the inner CV cycle's parameter optimization is finished by separating the training data into a (inner) test and (inner) training set once more. On the outer test fold, the optimized models can be checked for generalizability. This so-called layered CV method maximizes generalizability and has become the gold standard in psychiatric research.

A complete proposed system is shown in the following figure that illustrates the planned proposal explained in this chapter.

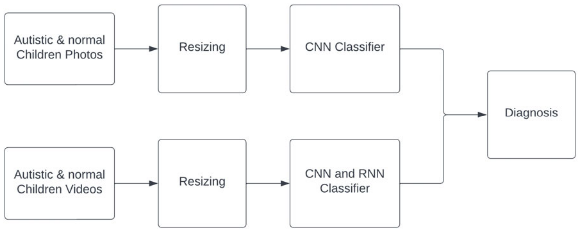

Autistic & normal Children Photos
Resizing
CNN Classifier
Diagnosis
Autistic & normal Children Videos
Resizing
CNN and RNN Classifier

Fig. 5 Complete Predictive system to diagnose Autism in children

### 3.1 Deep Learning

Deep learning is a machine learning and artificial intelligence (AI) technology that mimics human learning. Data science, which also encompasses statistics and predictive modelling, contains deep learning. Deep learning is particularly useful for data scientists who need to collect, analyze, and interpret vast amounts of data since it speeds up and simplifies the process. Deep learning may be used to automate prediction analysis at the most basic level.

In traditional machine learning, the learning process is supervised, and the programmer must be extremely precise when teaching the computer what sorts of things it should search for to decide whether a picture contains a dog. The computer's success rate is entirely based on the programmer's ability to correctly specify a feature set for a dog. Deep learning has the advantage of constructing the feature set without supervision. When compared to supervised learning, unsupervised learning is typically more accurate.

To achieve satisfactory accuracy, deep learning systems require a huge amount of data (Big Data) that can only be processed using cloud computing. Because it can generate intricate statistical models directly from its own repeating output, deep learning programming may develop accurate prediction models from massive volumes of unlabeled, unstructured data.

A variety of methodologies may be utilized to develop robust deep learning models. These tactics include learning rate degradation, transfer learning, starting from scratch, and dropping out.

The learning rate is a hyperparameter that determines how much the model changes when the weights are changed in response to the predicted mistake. Prior to the learning process, it is a number that defines the system or provides the conditions for its operation. When learning rates are excessively high, unstable training processes or the acquisition of a bad set of weights might occur. Too slow learning rates might lead to a protracted training process that becomes locked.

The learning rate decay method, also known as adaptable learning rates or learning rate annealing, is a strategy for improving performance while minimizing training time. The simplest and most common learning rate changes throughout training are techniques to lessen the learning rate over time.

Transfer learning is an approach that involves access to a network's internals and entails fine-tuning a previously trained model. To begin, users add fresh data to the network, including previously unknown categories. Once the network has been updated, new jobs with more specific classification skills can be completed. This approach has the benefit of requiring far less data than others, reducing calculation time to minutes or hours.

This approach requires the collecting of a large, labelled data set as well as the setup of a network architecture capable of learning the features and model. This strategy is especially useful for new apps and those with a lot of different output types. However, because it demands an exorbitant quantity of data, it is a less common method in general, causing training to take days or weeks.

Dropout is a strategy for avoiding overfitting in networks with numerous parameters by randomly eliminating units and their connections from the neural network during training. In disciplines including document classification, speech recognition, and computational biology, the dropout method has been proven to increase neural network performance on supervised learning tasks.

### 3.2 Convolutional Neural Network CNN

A Convolutional Neural Network CNN is a Deep Learning system that can take an image as input, give priority, and discriminate between various aspects in the picture. A CNN requires far less pre-processing than conventional classification methods. While simple techniques need hand-engineering of filters, CNN can learn these filters/characteristics with enough training.

The design of a CNN is based on the organization of the Visual Cortex and is similar to the neuron connection pattern in the human brain. Individual neurons can only respond to stimuli in the Receptive Field, a tiny portion of the visual field. If a set of comparable fields overlap, the entire visual region is covered.

A CNN can successfully capture the temporal and spatial relationships in an image with the right filters. The architecture achieves better fitting to the picture dataset because to the reduced number of parameters and reusability of weights. In other words, the network may be taught to recognize the image's level of complexity. As illustrated in Figure 4, the three major types of layers are convolution, pooling, and fully connected neural networks.

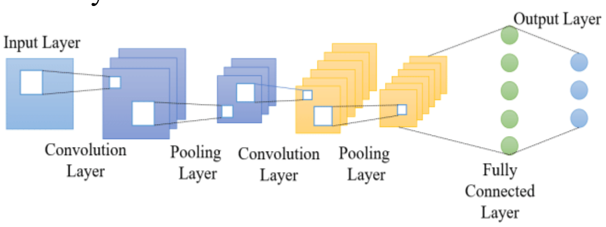


Fig. 6 Convolutional Neural Network Architecture.

A Convolution Operation is used to extract high-level properties such as edges from an input picture. There is no requirement for CNNs to have only one Convolutional Layer. The first Convolution Layer is usually in charge of collecting Low-Level data such as edges, color, gradient direction, and so on. The design responds to the High-Level properties as well with the addition of layers, giving us a network that knows the photos in the dataset as well as a person.

The procedure yields two sorts of results: one in which the convolved feature's dimensionality is reduced when compared to the input, and another in which the dimensionality is raised or unaltered. In the first situation, Valid Padding is used, and in the second case, Same Padding is used.

The Pooling layer is responsible for reducing the spatial size of the Convolved Feature. The computer power required to process the data is lowered because of dimensionality reduction. It also helps maintain the model's training process going smoothly by extracting rotational and positional invariant dominating features.

There are two forms of pooling: maximal pooling and average pooling. Max Pooling returns the maximum value from the region of the picture covered by the Kernel. On the other hand, Average Pooling returns the average of all the values from the image's Kernel section.

Noise can also be reduced by using Max Pooling. It de-noises and reduces the dimensionality of the data by removing all noisy activations. Average Pooling, on the other hand, is a dimensionality reduction strategy that uses noise suppression. As a result, we may conclude that Max Pooling outperforms Average Pooling.

### 3.3 Recurrent Neural Network

A recurrent neural network (RNN) is an artificial neural network that works with time series or sequence data. Ordinary feed forward neural networks are made to deal with data that is unconnected to one another. However, if we have data in a sequence where one data point is reliant on the previous data point, the neural network must be changed to account for these dependencies. RNNs have a concept of ‘memory,' which allows them to save the states or information of previous inputs in order to build the sequence's next output.

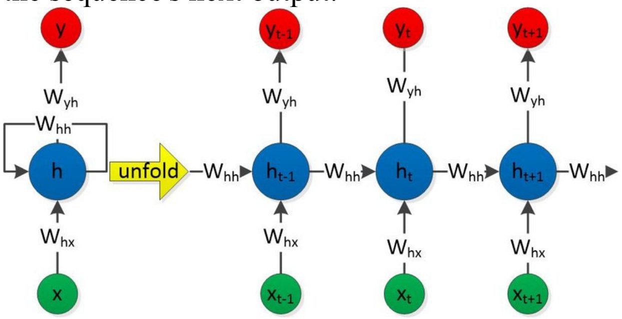


Fig. 7 Recurrent Neural Network Architecture.

As demonstrated in the first schematic of RNN as shown in figure 5, a simple RNN has a feedback loop. The grey rectangle represents a feedback loop that can be unrolled in three-time steps to form the second network seen above. Of course, the architecture can be changed so that the network unfolds in k time steps.

As a result, in an RNN's feedforward pass, the network computes the hidden unit values and output after k time steps. The network's weights are shared in a time-based manner. There are two sets of weights in each recurrent layer: one for the input and the other for the hidden unit. The final feedforward layer, which computes the final output for the kth time step, is like a classic feedforward network's ordinary layer.

RNNs have several advantages, including:

- The ability to handle sequence data and inputs of varied durations.
- The ability to save or "memorize" historical data.

The following are some disadvantages:

- The computation can be exceedingly slow.
- When making judgments, the network does not consider future inputs.
- The disappearing gradient problem occurs when the gradients used to compute the weight update approach zero, stopping the network from learning new weights. This difficulty becomes more apparent as the network grows deeper.

### 3.4 Autistic and Normal Children Images Classifier

That is the first module that has been used to diagnose the social interaction symptoms and determine whether they are for an autistic or normal child. We start explaining that model with the dataset used followed by the applied algorithms and their hyperparameters choice.

### 3.5 Dataset for autistic and normal images

A complete dataset that is composed of two folders containing photos for autistic children totaling 1032 photos while the other folder contains normal children photos totaling 1023 photos [9]. The photos need to be processed to crop the face of the child for the model to achieve better performance to determine the right child condition. The photos spans different children’s races and origins to better generalize to a wider range of children. The photos are varied in size. Therefore, they need to be resized before to be used for training and testing the classifier. Few photos’ samples are illustrated in Figure 6 with photos of both autistic and normal children derived from the dataset used for training and testing the predictive model

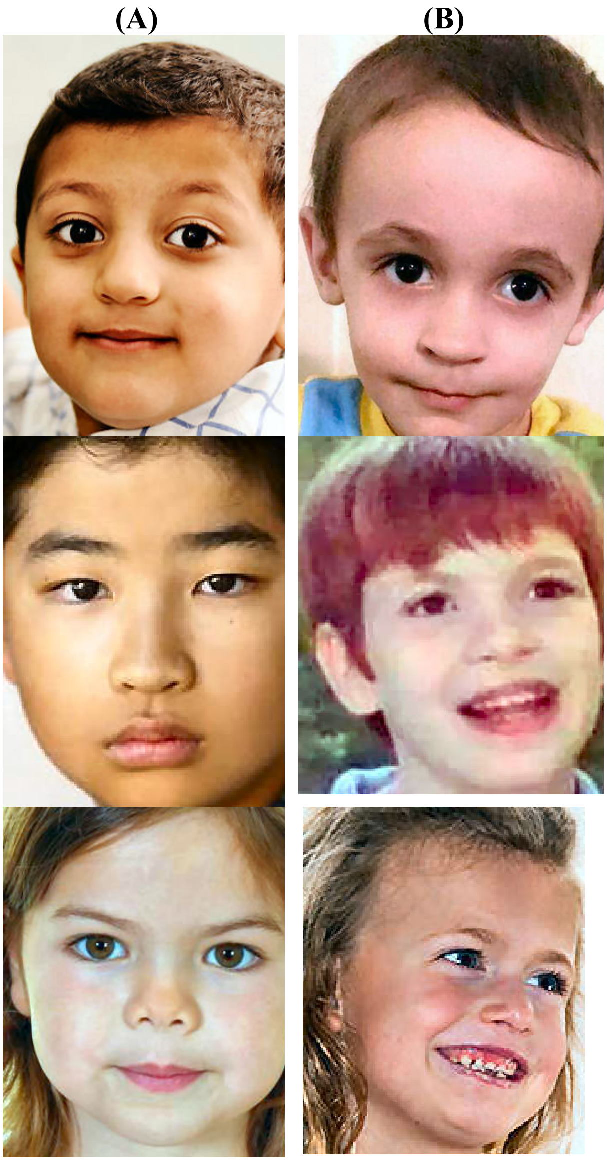


Fig. 8 Sample photos for (a) normal, (b) autistic children

### 3.6 Face segmentation

The Dlib library is undoubtedly one of the most widely used face recognition libraries. Face recognition is a Python package that encapsulates dlib's facial recognition routines into a simple, easy-to-use API. There are no parameters for the get frontal face detector function. A call to it returns the dlib library's pre-trained HOG + Linear SVM face detector. The HOG + Linear SVM face detection from Dlib is quick and accurate. The Histogram of Oriented Gradients (HOG) descriptor is not invariant to changes in rotation and viewing angle due to the way it works. That techniques will be applied to all children photos to segment their faces for better analysis using the CNN classifier.

### 3.7 Image Preparation:

The next step is to prepare the images in terms of resolution to be suitable for further processing using the CNN. All the images were scaled to be 224,224 while data were split between 80% for the training dataset and 20% for the testing dataset.

### 3.8 The Classification step:

In that module we have applied VGG16 using TensorFlow and using SGD optimization to find the CNN parameters. The input to the CNN layer is a 224 X 224 RGB picture with a fixed size. Each convolutional layer has an incredibly narrow receptive field: 33 (the smallest size that captures the ideas of left/right, up/down, and centre). In one of the settings, it additionally includes 11 convolution filters, which may be regarded of as a linear change of the input channels (followed by non-linearity). For 33 convolution layers, the convolution stride is set to 1 pixel, and the spatial padding of convolution layer input is set to 1 pixel to maintain spatial resolution after convolution.

To conduct spatial pooling, five max-pooling layers follow part of the convolution layers (not all the conv. layers are followed by max-pooling). Stride 2 is used to max-pool over a 22-pixel frame.

Three Fully Connected layers are placed following a stack of convolutional layers, the first two have a number of channels of 4096, while the third performs 1000-way ILSVRC classification and so has 1000 channels, one for each class. The final layer is the soft-max layer. The fully linked tiers in all networks are set up in the same way. All hidden layers with one output that has a sigmoid activation function employ the rectification (ReLU) activation function. Figure 7 depicts the preferred architecture.

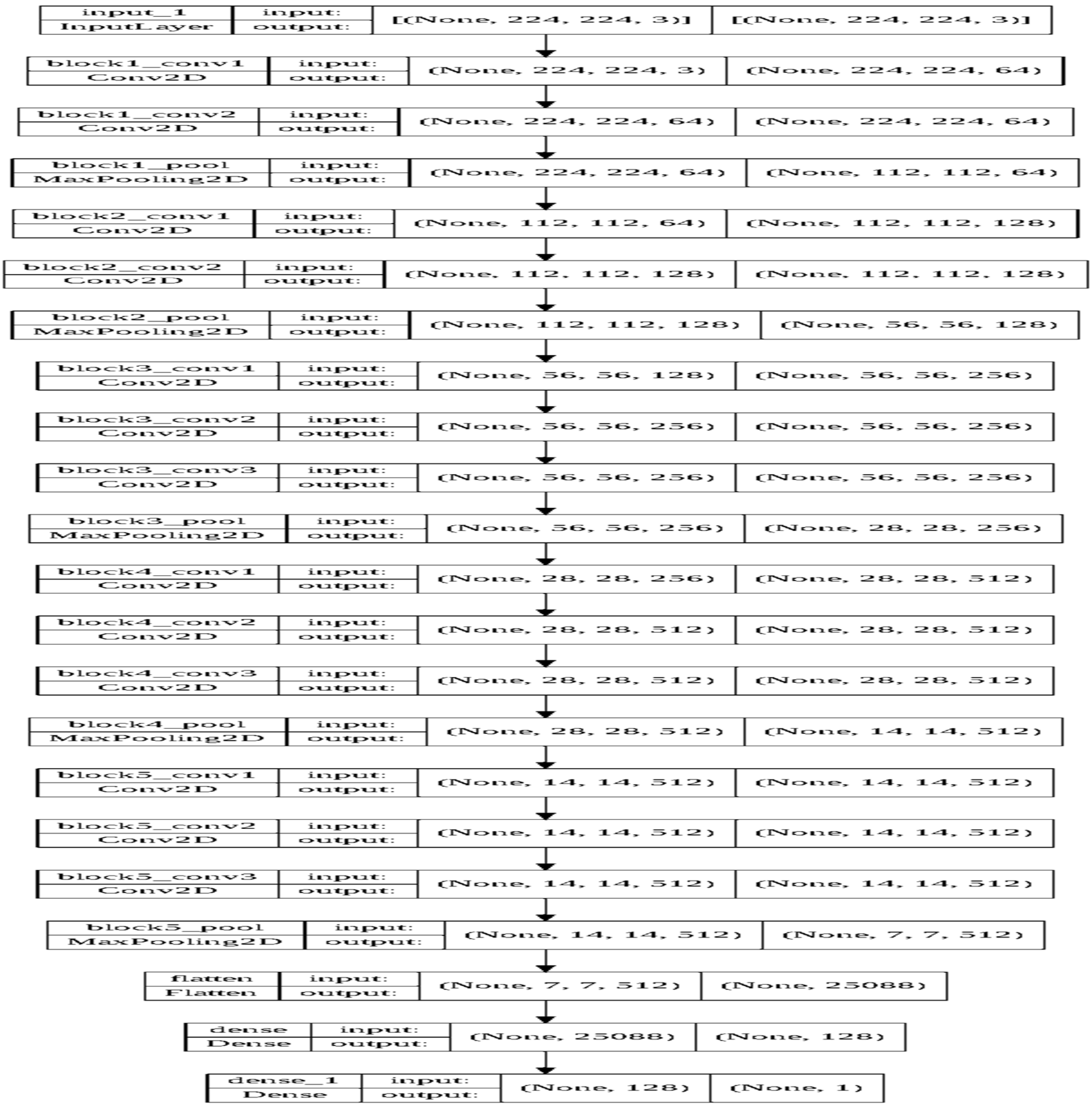

input_1 InputLayer input: output: [(None, 224, 224, 3)] [(None, 224, 224, 3)]
block1_conv1 Conv2D input: output: (None, 224, 224, 3) (None, 224, 224, 64)
block1_conv2 Conv2D input: output: (None, 224, 224, 64) (None, 224, 224, 64)
block1_pool MaxPooling2D input: output: (None, 224, 224, 64) (None, 112, 112, 64)
block2_conv1 Conv2D input: output: (None, 112, 112, 64) (None, 112, 112, 128)
block2_conv2 Conv2D input: output: (None, 112, 112, 128) (None, 112, 112, 128)
block2_pool MaxPooling2D input: output: (None, 112, 112, 128) (None, 56, 56, 128)
block3_conv1 Conv2D input: output: (None, 56, 56, 128) (None, 56, 56, 256)
block3_conv2 Conv2D input: output: (None, 56, 56, 256) (None, 56, 56, 256)
block3_conv3 Conv2D input: output: (None, 56, 56, 256) (None, 56, 56, 256)
block3_pool MaxPooling2D input: output: (None, 56, 56, 256) (None, 28, 28, 256)
block4_conv1 Conv2D input: output: (None, 28, 28, 256) (None, 28, 28, 512)
block4_conv2 Conv2D input: output: (None, 28, 28, 512) (None, 28, 28, 512)
block4_conv3 Conv2D input: output: (None, 28, 28, 512) (None, 28, 28, 512)
block4_pool MaxPooling2D input: output: (None, 28, 28, 512) (None, 14, 14, 512)
block5_conv1 Conv2D input: output: (None, 14, 14, 512) (None, 14, 14, 512)
block5_conv2 Conv2D input: output: (None, 14, 14, 512) (None, 14, 14, 512)
block5_conv3 Conv2D input: output: (None, 14, 14, 512) (None, 14, 14, 512)
block5_pool MaxPooling2D input: output: (None, 14, 14, 512) (None, 7, 7, 512)
flatten Flatten input: output: (None, 7, 7, 512) (None, 25088)
dense Dense input: output: (None, 25088) (None, 128)
dense_1 Dense input: output: (None, 128) (None, 1)

Fig.9 The chosen CNN architecture (VGG16).

### 3.9 Autistic and Normal Children Videos Classifier

That is the second module that has been used to diagnose the social interaction symptoms recorded in videos and determine whether they are for an autistic or normal child. We will describe the dataset and the Deep Learning neural network architecture used in this module along with the chosen hyperparameters.

### 3.10 Dataset for autistic and normal videos

That dataset was collected from different resources as the videos for autistic children were downloaded from (Self-Stimulatory Behaviors in the Wild for Autism Diagnosis Dataset – Roland Goecke) associated with a relevant paper that diagnosis different conditions of autism [10]. The total number of autistic videos is 75 divided between three types of symptoms namely, arm flapping, spinning, and head banging. We had to download those video clips and convert them to avi format suitable for further processing. To acquire videos suitable for the social interaction of normal children we have collected 100 videos for normal children and converted them into avi format. The two groups of videos will be processed further to be suitable for the CNN and RNN classifier training and testing.

### 3.11 Videos Preparation:

The videos needed to be resized to be suitable for the CNN we have chosen to extract the features of each video frame before those features to be fed in sequence to the Recurrent Neural Network. The videos resolution was scaled to be 224x224. The videos totaled 208 videos. Those videos were split between 90% for training and 10% for testing.

### 3.12 The deep learning classifier

#### Inception v3 for features extraction

The used classifier consisted of two deep learning neural networks in sequence. The first one is the inception v3 which has greater model adaption as it uses numerous strategies to optimize the network.

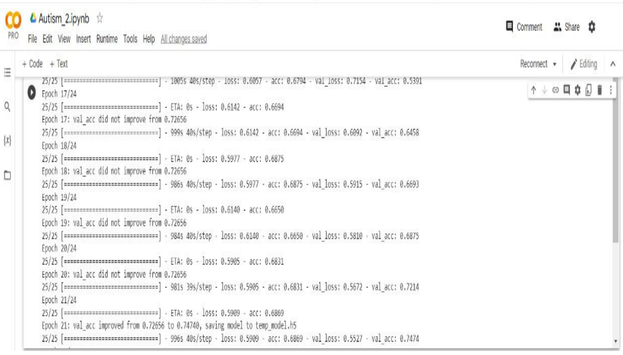
It is more efficient and has a more extensive network than the Inception V1 and V2 models, but its speed is unaffected. Moreover, it has less computational cost as well as employing auxiliary Classifiers for regularization. The final layer will serve as a feature vector with size 2048 that will be fed to the Recurrent Neural Network.

#### Gated Recurrent Unit

GRUs are a more advanced variant of the recurrent neural network. GRU employs the so-called update gate and reset gate to overcome the vanishing gradient problem of a regular RNN. In essence, there are two vectors that determine what data should be sent to the output. They are unique in that they can be trained to retain knowledge from the past without having to wash it away over time or delete information that is unrelated to the forecast. We have used 8 Gated Recurrent Units to classify the videos.

Each unit consists of Update gate, reset gate, and current memory content. The memory allows to ca pture the inherent sequence of frames in the video to predict whether the social interaction it captures the association of autism or normal behavior.

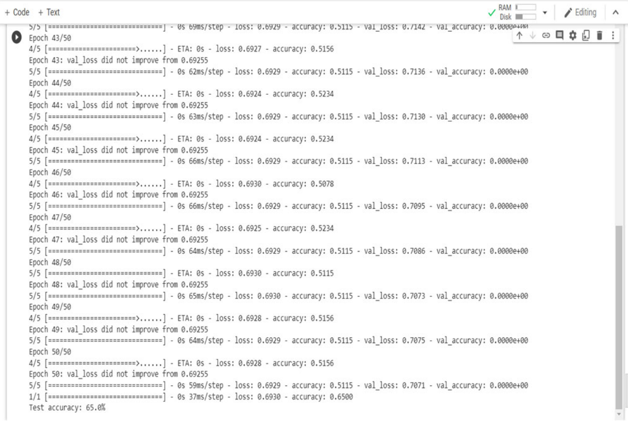
## 4. Results:



```
25/25 [==============================] - 1005s 40s/step - loss: 0.6057 - acc: 0.6794 - val_loss: 0.7154 - val_acc: 0.5391
Epoch 17/24
25/25 [==============================] - ETA: 0s - loss: 0.6142 - acc: 0.6694
Epoch 17: val_acc did not improve from 0.72656
25/25 [==============================] - 999s 40s/step - loss: 0.6142 - acc: 0.6694 - val_loss: 0.6092 - val_acc: 0.6458
Epoch 18/24
25/25 [==============================] - ETA: 0s - loss: 0.5977 - acc: 0.6875
Epoch 18: val_acc did not improve from 0.72656
25/25 [==============================] - 986s 40s/step - loss: 0.5977 - acc: 0.6875 - val_loss: 0.5915 - val_acc: 0.6693
Epoch 19/24
25/25 [==============================] - ETA: 0s - loss: 0.6140 - acc: 0.6650
Epoch 19: val_acc did not improve from 0.72656
25/25 [==============================] - 984s 40s/step - loss: 0.6140 - acc: 0.6650 - val_loss: 0.5810 - val_acc: 0.6875
Epoch 20/24
25/25 [==============================] - ETA: 0s - loss: 0.5905 - acc: 0.6831
Epoch 20: val_acc did not improve from 0.72656
25/25 [==============================] - 981s 39s/step - loss: 0.5905 - acc: 0.6831 - val_loss: 0.5672 - val_acc: 0.7214
Epoch 21/24
25/25 [==============================] - ETA: 0s - loss: 0.5909 - acc: 0.6869
Epoch 21: val_acc improved from 0.72656 to 0.74740, saving model to temp_model.h5
25/25 [==============================] - 996s 40s/step - loss: 0.5909 - acc: 0.6869 - val_loss: 0.5527 - val_acc: 0.7474
```

Fig. 10 vad_acc photo model

In the image model, we have reached an accuracy of 74.74%, and this accuracy is acceptable compared to previous research in the field of autism spectrum.



```
5/5 [==============================] - 0s 69ms/step - loss: 0.6929 - accuracy: 0.5115 - val_loss: 0.7142 - val_accuracy: 0.0000e+00
Epoch 43/50
4/5 [========================>......] - ETA: 0s - loss: 0.6927 - accuracy: 0.5156
Epoch 43: val_loss did not improve from 0.69255
5/5 [==============================] - 0s 62ms/step - loss: 0.6929 - accuracy: 0.5115 - val_loss: 0.7136 - val_accuracy: 0.0000e+00
Epoch 44/50
4/5 [========================>......] - ETA: 0s - loss: 0.6924 - accuracy: 0.5234
Epoch 44: val_loss did not improve from 0.69255
5/5 [==============================] - 0s 63ms/step - loss: 0.6929 - accuracy: 0.5115 - val_loss: 0.7130 - val_accuracy: 0.0000e+00
Epoch 45/50
4/5 [========================>......] - ETA: 0s - loss: 0.6924 - accuracy: 0.5234
Epoch 45: val_loss did not improve from 0.69255
5/5 [==============================] - 0s 66ms/step - loss: 0.6929 - accuracy: 0.5115 - val_loss: 0.7113 - val_accuracy: 0.0000e+00
Epoch 46/50
4/5 [========================>......] - ETA: 0s - loss: 0.6930 - accuracy: 0.5078
Epoch 46: val_loss did not improve from 0.69255
5/5 [==============================] - 0s 66ms/step - loss: 0.6929 - accuracy: 0.5115 - val_loss: 0.7095 - val_accuracy: 0.0000e+00
Epoch 47/50
4/5 [========================>......] - ETA: 0s - loss: 0.6925 - accuracy: 0.5234
Epoch 47: val_loss did not improve from 0.69255
5/5 [==============================] - 0s 64ms/step - loss: 0.6929 - accuracy: 0.5115 - val_loss: 0.7086 - val_accuracy: 0.0000e+00
Epoch 48/50
5/5 [==============================] - ETA: 0s - loss: 0.6930 - accuracy: 0.5115
Epoch 48: val_loss did not improve from 0.69255
5/5 [==============================] - 0s 65ms/step - loss: 0.6930 - accuracy: 0.5115 - val_loss: 0.7073 - val_accuracy: 0.0000e+00
Epoch 49/50
4/5 [========================>......] - ETA: 0s - loss: 0.6928 - accuracy: 0.5156
Epoch 49: val_loss did not improve from 0.69255
5/5 [==============================] - 0s 64ms/step - loss: 0.6929 - accuracy: 0.5115 - val_loss: 0.7075 - val_accuracy: 0.0000e+00
Epoch 50/50
4/5 [========================>......] - ETA: 0s - loss: 0.6928 - accuracy: 0.5156
Epoch 50: val_loss did not improve from 0.69255
5/5 [==============================] - 0s 59ms/step - loss: 0.6929 - accuracy: 0.5115 - val_loss: 0.7071 - val_accuracy: 0.0000e+00
1/1 [==============================] - 0s 37ms/step - loss: 0.6930 - accuracy: 0.6500
Test accuracy: 65.0%
```

Fig. 11 test acc video model

We have reached an accuracy of 65% for the video sample, and this accuracy is high given the privacy of the data of the target group in the research.

## 5. Conclusions

That chapter discussed the steps we are going to perform on the selected datasets that included both autistic and normal photos and videos. Several machine learning techniques will be used including CNN and RNN to process both the images and the videos. The combination between CNN and RNN was preferred because of the nature of the videos and the need to capture the social interaction which can be revealed from the sequence of frames composing the videos. However, for photos CNN was enough to capture the facial features. The use of both photos and videos was suggested to increase the performance of the predictive model.

## References

[1] Xiaoyi Yu, Lingyi Wu, Qingfeng Liu and Han Zhou, "Children tantrum behaviour analysis based on Kinect sensor," 2011 Third Chinese Conference on Intelligent Visual Surveillance, 2011, pp. 49-52, doi: 10.1109/IVSurv.2011.6157022.

[2] P. Rani, "Emotion Detection of Autistic Children Using Image Processing," 2019 Fifth International Conference on Image Information Processing (ICIIP), 2019, pp. 532-535, doi: 10.1109/ICIIP47207.2019.8985706.

[3] S. Liang, A. Q. M. Sabri, F. Alnajjar and C. K. Loo, "Autism Spectrum Self-Stimulatory Behaviors Classification Using Explainable Temporal Coherency Deep Features and SVM Classifier," in IEEE Access, vol. 9, pp. 34264-34275, 2021, doi: 10.1109/ACCESS.2021.3061455.

[4] S. K. Jarraya, M. Masmoudi and M. Hammami, "Compound Emotion Recognition of Autistic Children During Meltdown Crisis Based on Deep Spatio-Temporal Analysis of Facial Geometric Features," in IEEE Access, vol. 8, pp. 69311-69326, 2020, doi: 10.1109/ACCESS.2020.2986654.

[5] M. Kokot, F. Petric, M. Cepanec, D. Miklić, I. Bejić and Z. Kovačić, "Classification of Child Vocal Behavior for a Robot-Assisted Autism Diagnostic Protocol," 2018 26th Mediterranean Conference on Control and Automation (MED), 2018, pp. 1-6, doi: 10.1109/MED.2018.8443030.

[6] Fei Dan, Lin Zhao, Changqing Suo and Qianqian Sun, "An experiment of influence of sports games on 3–6 years old young children's social behaviors," 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), 2011, pp. 5397-5400, doi: 10.1109/AIMSEC.2011.6009799.

[7] Y. Yang et al., "Emotional and behavioral problems, social competence and risk factors in 6–16-year-old students in Beijing, China", PLOS ONE, vol. 14, no. 10, p. e0223970, 2019. Available: 10.1371/journal.pone.0223970 [Accessed 18 December 2021].

[8] X, Dongxin, Jill Gilkerson, and Jeffrey A. Richards. "Objective child behavior measurement with naturalistic daylong audio recording and its application to autism identification." 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2012.

[9] GitHub. 2022. GitHub - mm909/Kaggle-Autism: Detecting Autism Spectrum Disorder in Children With Computer Vision - Adapting facial recognition models to detect Autism Spectrum Disorder. [online] Available at: <https://github.com/mm909/Kaggle-Autism> [Accessed 15 May 2022].

[10] Roland Goecke. 2022. Self-Stimulatory Behaviours in the Wild for Autism Diagnosis Dataset. [online] Available at: <https://rolandgoecke.net/research/datasets/ssbd/> [Accessed 15 May 2022].

**Abdulrahman Turki Alharbi** received the B.S. in science of mathematics from University of Hafar Al Batin and M.S. degrees in data science, degrees in data science. in 2020 and 2022, respectively. His research interest includes data processing, modeling and data analysis , Business Analytics , artificial intelligence .

**Hadi Mohmmad Alyami** received the B.S in Information Technology . and M.S. degrees in data science, from University of Tabuk. in 2020 and 2022, respectively. His research interest includes data processing , modeling and data analysis , deep learning .

**Saleh Eid Alenazi** received the B.S. in Computer Science and M.S. degrees in data science from University of Tabuk . in 2019 and 2022, respectively. His research interest includes data processing, modeling and data analysis , encryption , deep learning , privacy .

**Saud Awadhallah Alharbi** received the B.S. in information of science, from University of Taibah and M.S degrees, from in data of science from University of Tabuk . in 2020 and 2022, respectively. His research interest includes data processing, Data Visualization and data analysis , Business Analytics. , artificial intelligence .

**Zaid Bassfar** received the B.S. in Computer Science in 2007, and M.S. degrees in Information Technology and Communication in 2010. He received the PhD degree in Web Applications in 2014. He has been an assistant professor since 2014 at College of computer and Information Technology at University of Tabuk. His research interest includes Web Applications – Virtual Reality – Emerging Technology - Virtual Learning - E-learning – M-learning – Multimedia – Human Computer Interaction.