Biofeedback Upper Limb Assessment Using Electroencephalogram, Electromyographic and Electrocardiographic with Machine Learning in Signal Classification

Kai Liang Lew, Kok Swee Sim, Shing Chiang Tan and Fazly Salleh Abas

Abstract—**Physical disability or arm paralysis is a common symptom for the post-stroke survivor. The upper limb rehabilitation is introduced to improve the motor ability of the upper limb and recovery from stroke. However, the recovery rate of the motor ability upper limb is based on physical condition and therapy performance of a patient (subject). The rehabilitation may require more manpower at a center and is time consuming for a physiotherapist to monitor a patient during rehabilitation without the use of technology. The purpose of research in this paper is to evaluate the condition of subjects using a deep learning model with biosignal devices after virtual reality (VR) upper limb assessment. Fifteen control persons and fifteen post-stroke patients have performed two games under VR upper limb assessment, namely, Touch the Ball, and Stack the Cube. The patients were equipped with an electroencephalogram (EEG), electromyographic (EMG), and electrocardiographic (ECG). The measurements were taken before, during, and after the assessment. A common practice in data handling is that all EEG, EMG and ECG signals are pre-processed to remove noises or to condition data. In this work, all three raw biosignals were collectively represented as images. They were used to train deep learning models (namely, convolutional neural network and long-short term memory) of which the models were used to evaluate the condition of a subject. The classification performance of the deep neural network in classifying the biosignals is highly accurate and precise.**

Index Terms— **Virtual Reality, Rehabilitation, Biosignals, Unreal Engine 4**

I. Introduction

Stroke is a major threat in causing physical disability worldwide. It leads to several impairments such as paralysis, motor dysfunction, and weakness. Upper limb motor impairment can affect daily life activities and perceived quality of life. Therefore, rehabilitation is introduced for improving and assisting stroke patients to improve their motor function and life quality. During rehabilitation, it is important to monitor the biosignals of a patient to determine stroke symptoms and the patient's stroke status.

Biosignals are generated by the human body with different wearable sensors. Biosignals represent a portion of the body's signal and one sensor can only measure a biosignal at a time. It is inadequate to evaluate a patient's condition before, during, and after the rehabilitation based only on a single source of biosignal. Therefore, more than one sensor could be used in many rehabilitation programs to obtain more information about the patient. The most commonly used biosignal sensors during rehabilitation are EEG, EMG, and ECG. EEG measures the voltage emitted from the scalp due to the ionic current flows within the neurons in the brain [1]. The muscle signal can be measured by using EMG while moving the hand. ECG is used to measure the heartbeat signal.

Biofeedback systems have been used in the field of rehabilitation to facilitate [2] a subject to restore motor function after an injury [3]. It provides biological information of patients in real-time and it usually involves measurement of a target biomedical variable and relaying it to the user such as direct feedback and transformed feedback [3]. It may offer the opportunity to improve accuracy during functional tasks, increase patient engagement in rehabilitation, and reduce frequency of making a physical appointment with healthcare professionals to update rehabilitation progress.

General signal classification can be a difficult task when classifying non-pre-processed signals. This is due to the signals without pre-processing may consist of various factors such as interference, low SNR, fading, phase, and frequency offsets distort the received signal which can affect overall results and classification [4]. However, signal classification can be performed easily by using machine learning or deep learning. The use of machine learning and deep learning in signal classification is popular in neuroscience.

In this research, a subject will wear biosignal devices before performing two upper limb assessment activities. The biosignals (ECG, EMG and EEG) are collected through an Arduino board and visual studio. The raw signals are plotted graphically and saved as an image using Python. Then, deep learning models are trained with a collection of images before

Manuscript received January 11, 2021; revised January 5, 2022. This work was supported in part by the TM R&D under Grant (MMUE/180026) for the research program of Rehabilitation Using Biofeedback System.

K.L. Lew is a postgraduate student from Faculty of Engineering & Technology, Multimedia University, 75450 Melaka, Malaysia. (phone: 6012-6491832; e-mail: lewkailiang@gmail.com).

K.S. Sim, the corresponding author, is a Professor from Faculty of Engineering & Technology, Multimedia University, 75450 Melaka, Malaysia. (e-mail: kssim@mmu.edu.my).

S.C. Tan is a professor from Faculty of Information Science and Technology, Multimedia University, 75450 Melaka, Malaysia. (e-mail: sctan@mmu.edu.my)

F. S. Abas is an associate professor from Faculty of Engineering & Technology, Multimedia University, 75450 Melaka, Malaysia. (e-mail: fazly.salleh.abas@mmu.edu.my).

classifying the signals (images) without using signal processing techniques. The objective of this research is to evaluate the condition of the subjects (by referring to biosignals being captured in images) after finishing the VR upper limb assessment.

The paper is organised as follows: Section 2 describes all related work to this research. Section 3 explains the procedure of the proposed work and deep learning architectures in classification. Section 4 shows the result of the assessment and classification. Section 5 presents a discussion on the results. Lastly, section 6 concludes the work and draws some recommendations for further work.

II. Related work

Sim et al. [5] designed two VR-based applications for rehabilitation by using a leap motion sensor, Microsoft Kinect and Mobile-based VR headset. The first application was a finger VR-based rehabilitation system for treating the post-stroke patient. The patient was required to wear a headset which was attached to a leap motion sensor before moving fingers. The first game was Pick and Place for post-stroke patients. The game was created by using the Unity game engine and allowed the participant to make finger movement and gripping. The participant was required to pick up a block and stack them up accordingly in the VR environment

Marin-Pardo et al. [6] presented that feedback of muscle activity measured using surface EMG was more effective than EEG biofeedback. They used a VR training software namely REINVENT which was developed for stroke rehabilitation. It showed EEG biofeedback of brain activity through the recorded EMG signals. The EEG and EMG signals were pre-processed by using Matlab scripts such as DC-offset correction, full-wave rectification, and filtering. There were three participants tested on VR rehabilitation. They compared the success rate of EMG and EEG by using a paired-sample t-test between each modality for each participant. The result showed the participants could perform with a better motor ability when using EMG biofeedback as compared to EEG biofeedback.

Paula et al. [7] assessed the motor outcome in chronic stroke patients that underwent a robot-assisted rehabilitation program by using quantitative EEG (QEEG). Ten post-stroke patients in the chronic phase participated in the study. The patients interacted with the robot through the handle mounted as an end-effector. In the rehabilitation protocol, each patient was required to do twelve training sessions. Other than that, the resting state EEG recordings were performed before and after treatment. The EEG data was analysed in MATLAB after pre-processing the EEG data with several filters. The results showed the EEG data collected by using QEEG were useful in giving information for clinical decision-making and predicting the motor outcome.

Nagabushanam et al. [8] proposed two-layer Long Short Term Memory (LSTM) and 4-layers improved neural network (INN) deep learning algorithms to improve performance in classifying the EEG signals. A performance comparison between the deep learning method with logistic regression (LR) and support vector machine (SVM) was made in terms of accuracy, precision, recall and F1 score. The proposed method achieved 71% accuracy in INN and 78% accuracy in LSTM which were better than the performance of SVM and LR.

Based on the literature review, the first review [5] is related to VR headset with leap motion sensor. The limitation of the leap motion sensor is it can only work well under a low lighting room as it uses IR to detect hand movement. Therefore, HTC Vive devices can be used in any lighting condition in a room. Moreover, the hand movement can easily be detected by using the VR controller. In [6], EMG biofeedback rehabilitation and EEG biofeedback rehabilitation are compared to each other. In this comparison, 3 biosignal devices were used and signals collected from these devices could be referred to determine the status of the user. The third paper [7] uses the EEG device in rehabilitation and pre-processing it before predicting the outcome. In [8], an LSTM and an improved neural network were used to classify the EEG signals.

III. Methodology

A. Overview of Methodology

A participant was required to use HTC Vive to do 'Stack the Cube' (STC) and 'Touch the Ball' (TTB) training in the virtual world while wearing biosignal devices such as ECG, EMG, and EEG. The biosignals data were obtained and plotted through Python. Fig. 1 shows the block diagram of the proposed method in this study.

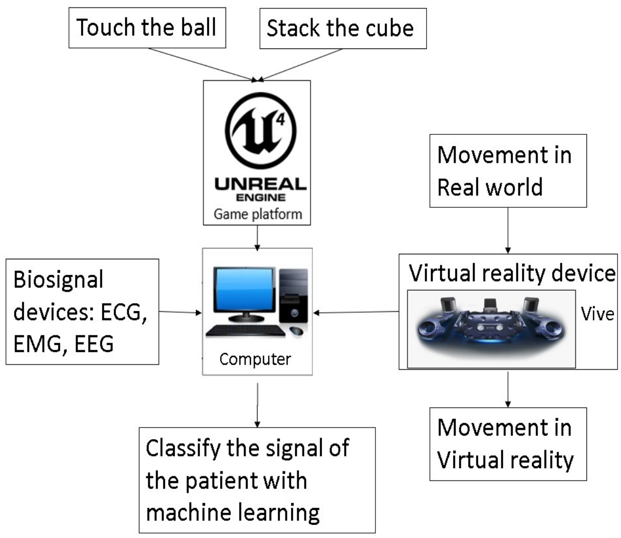

Fig. 1. Smart Biofeedback Virtual Reality Rehabilitation Method

The STC and TTB were designed to assess upper limb motor performance. Moreover, the patient's status can be monitored during and after the training. The ECG, EMG, and EEG signals obtained from the patient were mapped and saved as images. These images were used to train a hybrid deep learning model (i.e., CNN-LSTM) before identifying the biosignals from the image, either in a normal or stroke condition.

B. HTC Vive

The HTC Vive consists of a head-mounted display (HMD), two handheld controllers, and two base stations (sensors) [9]. The HTC Vive HMD virtualises virtual content and it has a resolution of 1080 pixels×1200 pixels per-eye and operates at 90 Hz. The HMD allows for 360-degree position tracking and is integrated with the audio system [10]. The handheld controller helps the user to contact the content in the virtual world through physical movement in the real world and it consists of a haptic sensor which simulates the sense of touch. It utilises the low-latency tracking technology to assess the headset's relative position. The controller performs several actions such as grab, hold, select options, etc in the virtual world. HTC Vive uses two laser emitters as sensors to perform tracking on the handheld controllers and HMD. The sensors alternatingly send out horizontal and vertical infrared laser sweeps spanning 120° in each direction [11]. Two sensors are placed in front and one sensor placed behind the user to perform a 360-degree position tracking. The application is run on an RTX 2080Ti desktop computer. Two requirements shall be fulfilled before using VR; first, the area must be at least 1 m long and 2 m wide; second, the area must be free from any obstacle. This is to avoid any unnecessary incidents when using the VR device.

C. Biosignal devices

Biosignals provide communication between biosystems and biosignal devices to measure the biosignal information produced by the electrical activity from the biological activity within different tissues and organs of the human body [12]. Three biosignal devices are used in this study namely electrocardiography (ECG), electroencephalogram (EEG), and electromyographic (EMG). The ECG device, SparkFun Single Lead Heart Rate Monitor-AD8232 is used. It is a cost-effective board that measures the electrical activity of the heart over a period of time. The ECG is generated from an electrical conduction mechanism that produces electrical impulses with a polarization and depolarization effect in the cardiac tissue. The ECG indicates life and the heartbeat of a person. The ECG signals are sampled at 1k Hz and after the collection is the low-pass filter, high-pass filter, band-pass filter, and rectifier.

The EMG device from OYMotion, Gravity Analog EMG Sensor is used. This sensor integrates a filter circuit and amplifier circuit. The ±1.5mV EMG signal could be amplified 1000 times and depressed noise, especially power frequency interference. The signal output is generated in an analog form, using 1.5 V as a reference voltage. The signal strength depends on the intensity of muscle activities. Whenever a muscle fiber contracts, a small electric current is generated, creating an EMG signal. During measurement, metal dry electrodes are connected to the device and attached to the patient's bicep and brachioradialis.

The EEG device used in this research is EMOTIV Insight 5 Channel Mobile Brainwear (Insight). This device is designed for self-quantification, brain-computer interface, and field research. It can sample up to 5 channels of EEG with a sampling rate of 2048 Hz. The sensor material used is semi-dry polymer and it has a built-in digital 5th order sinc filter. The Insight set up time is about 1-2 minutes and it can be connected wirelessly. In 1929, Hans Berger made the first recording on the human brain [13].

D. CNN-LSTM Model

In the proposed CNN-LSTM model, Convolution Neural Network (CNN) is applied to encode the image in a fixed-size feature representation, and Long Short Term Memory (LSTM) [14], is used to decode these features from the encoder. The attention mechanism passes its output as an additional input as LSTM and receives the output of LSTM at each timestep. It then outputs a weighted encoded image that is concatenated to the refined pixel of the encoded image to generate the LSTM input, which consists of one pixel of the original image and the encoded image. Fig. 2 shows the architecture of the CNN-LSTM Model

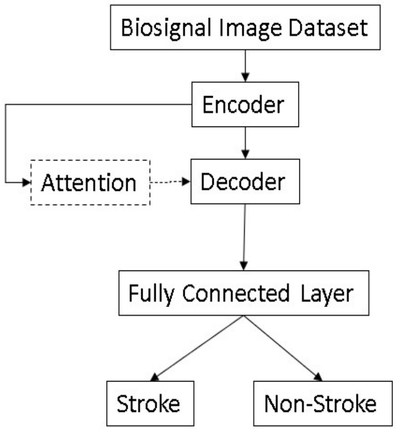

Fig. 2. The Architecture of CNN-LSTM with Attention

1) Encoder Layer

Deep residual network (ResNet) [15] may have different numbers of hidden layers. Examples of popular ResNets are ResNets18, which has 18 hidden layers; ResNet34, which has 34 hidden layers; and ResNet50, which has 50 hidden layers. ResNets apply average pooling on each channel and squeeze each feature map into a single numeric value. The ResNet has an advantage whereby the more the depth of layer, the higher the accuracy of the network model. Moreover, it can achieve a higher accuracy while avoiding the negative outcome. The ResNet solves the "vanishing gradient" problem by using "identifying shortcut connections" to identify layers that do nothing and skip them. These identical layers are skipped during the training process. When the identical layers are skipped, it will re-use the activation functions from the previous layers to reduce the network into only a few layers, which improves the learning speed. When the network is retrained, the same layers are expanded and help the network to explore more in the feature space. Fig. 3 shows the residual block of the deep residual network

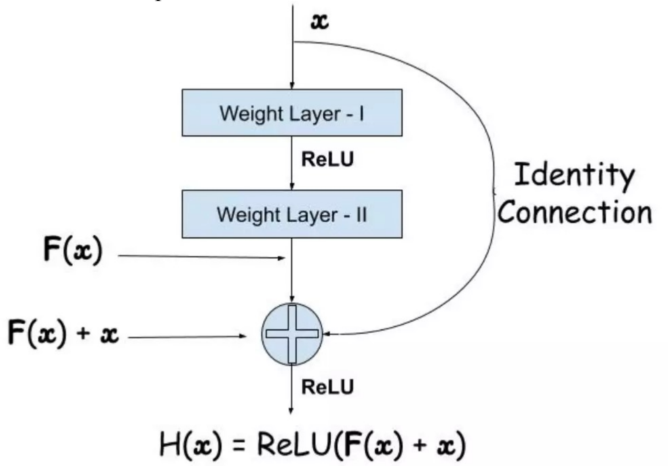

Fig. 3. A Residual Block of Deep Residual Network

For the activation function, a rectified linear unit (Relu) is used in a simple network. It is responsible for transforming the weighted summary data from the node into a node or output activation for that input. If it is positive, it will generate the input directly, else it will produce zero. Linear layer 1 equation is shown in (1) and equation for Relu operation on a linear layer 1 shown in (2). The linear layer 2 equation is shown in (3) and equation for Relu operation on a linear layer 2 shown in (4). Fig. 4 shows a block without skip connection and a block with skip connection. When the block has a skip connection, the Relu will be discarded because a shortcut is added before the Relu. Therefore, the Relu operation on linear layer 2 is re-written in (5).

$$z^{[l+1]}=W^{[l+1]}+a^{[l]}+b^{[l+1]} \tag{1}$$

$$a^{[l+1]}=g(z^{[l+1]}) \tag{2}$$

$$z^{[l+2]}=W^{[l+2]}+a^{[l]}+b^{[l+2]} \tag{3}$$

$$a^{[l+2]}=g(z^{[l+2]}) \tag{4}$$

$$a^{[l+2]}=g(z^{[l+2]}+a^{[l]}) \tag{5}$$

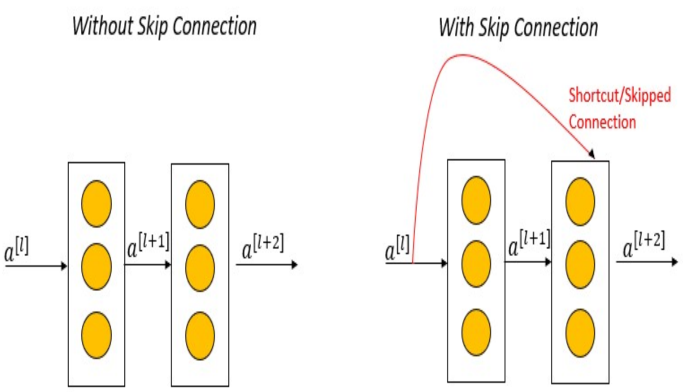

Fig. 4. A Block Without Skip Connection and With Skip Connection

The ResNets applied average pooling on each channel and squeezed each feature map into a single numeric value. The value will fit through a two-layer neural network for further processing and the fully connected layer is activated by a softmax function and outputs a vector of feature maps. Fig. 5 shows the example of ResNet architecture.

The encoder encodes the input image with 3 colour channels into a smaller encoded image. It summarises all the useful representations of the original image. The extractor can produce each of the D-dimensional of the L vector based on the part of the image. Equation 6 shows the extractor formula. The features are extracted from a lower convolutional layer to obtain the feature vectors and portions of the 2-D image which allows the decoder to select a subset of the feature vectors by focusing on certain parts of an image [16].

$$a=\{a_1,\ldots,a_L\},a_i \in R^D \tag{6}$$

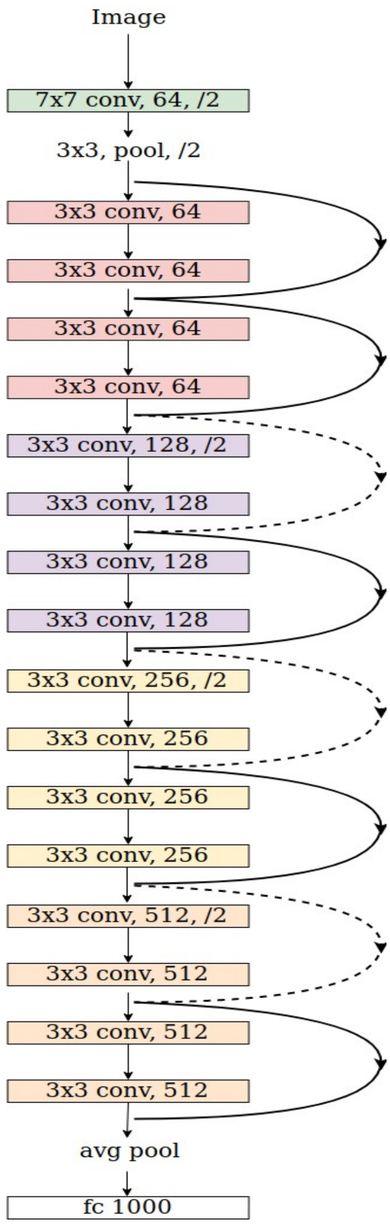

Fig. 5. Example of ResNet Architecture

In this study, the encoder uses convolution operation of ResNet to perform feature extraction. The ResNet is used for feature extraction. Therefore, the fully connected layer in the architecture is removed and the average pooling layer is replaced with adaptive averaging pooling. An adaptive two-dimension average pooling is applied to get a scalar value followed by two Relu activation functions. The architecture of ResNet used in the study is shown in Fig. 6

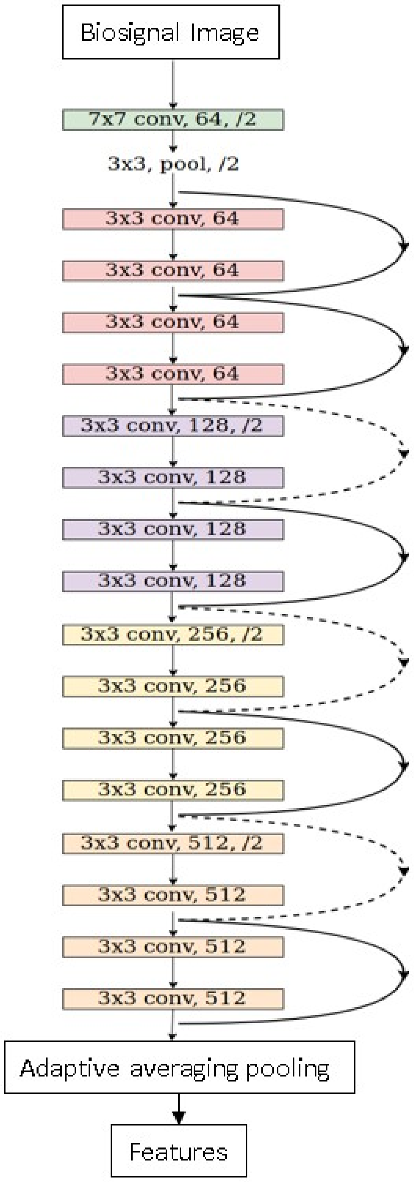

Fig. 6. ResNet architecture in the Study

2) Attention Layer

The feature obtained from the encoder will be passed to the attention layer to focus on the relevant feature. An attention function can be described as mapping a query and a set of key-value pairs to an output, where the query, keys, values, and output are all vectors [17]. The output is computed as a weighted sum of the values, where the weight assigned to each value is computed by a compatibility function of the query with the corresponding key. The attention mechanism combines the encoder output in a linear layer and previous decoder output in a linear layer and passes them through a non-linearity activation, and finally passes them through a softmax operation to represent the weights. The weights are then multiplied by the encoded image to generate the weighted encoded image. Mathematically, the attention mechanism can be described by (7), (8), (9), and (10). The output is a context vector, $\hat{c}_t$ which represents the weighted encoded image. The h_t is the previous decoder output, w_h, w_v are the linear layer weights. w_a is the softmax weights and $\hat{c}_t$ is the context vector generated and V_i is the encoded image features. The attention mechanism shown in Fig. 7

$$z_t = relu(h_t w_h + V_i w_v) \tag{7}$$

$$f_t = w_a z_t \tag{8}$$

$$\alpha = \frac{e^{f_t}}{\sum e^{f_t}} \tag{9}$$

$$\hat{c}_t = \sum_{i=1}^{pixels} \alpha_i V_{t_i} \tag{10}$$

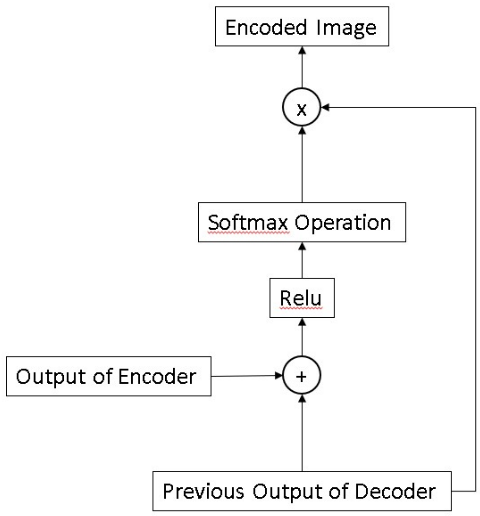

Fig. 7. The attention mechanism

3) Sequence-to-sequence (Seq2seq)

The Seq2seq model can map a variable input sequence to a variable-length output sequence using encoder-decoder. The encoder reads the input sequence and outputs a single vector [18]. The decoder reads the vector to produce an output sequence. The encoder of the seq2seq model used a Convolution Neural Network (CNN) and the decoder used Long Short Term Memory (LSTM) cell [19].

4) Decoder Layer

Long Short Term Memory (LSTM) (Fig. 8) has been chosen as the decoder. The decoder uses a pixel of the image and compares it with a pixel of the original image. There are self-attention layers in the decoder which allow each position in the decoder to attend to all positions in the decoder. This prevents leftward information flow in the decoder and preserves the auto-regressive property. Equations 11 to 16 are i, f, c, o, and h which refer to input, forget, memory, output, and hidden state.

$$i = \sigma(W_{ii}x + b_{ii} + W_{hi}h + b_{hi}) \tag{11}$$

$$f = \sigma(W_{if}x + b_{if} + W_{hf}h + b_{hf}) \tag{12}$$

$$g = \tanh(W_{ig}x + b_{ig} + W_{hg}h + b_{hg}) \tag{13}$$

$$o = \sigma(W_{io}x + b_{io} + W_{ho}h + b_{ho}) \tag{14}$$

$$c' = f*c + i*g \tag{15}$$

$$h' = o*\tanh(c') \tag{16}$$

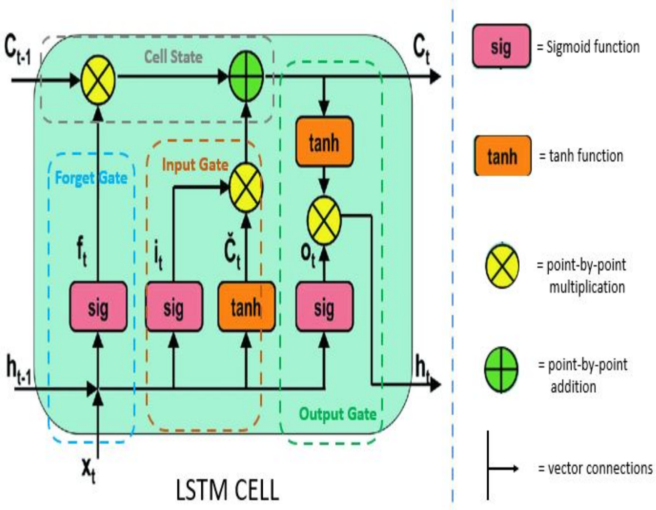

Fig. 8. Long Short Term Memory [20]

5) Fully Connected Layer

The fully connected layer is shown in Fig. 9. It can flatten the output of the decoder into a single vector of values to calculate the probability for each label based on type of model [21]. The EEG classification will classify the EEG signal and label the output as stroke or normal. The EMG classification will classify the EMG signal and label the output as stroke or normal. The ECG classification will classify the ECG signal and label the output as fast pace or slow pace.

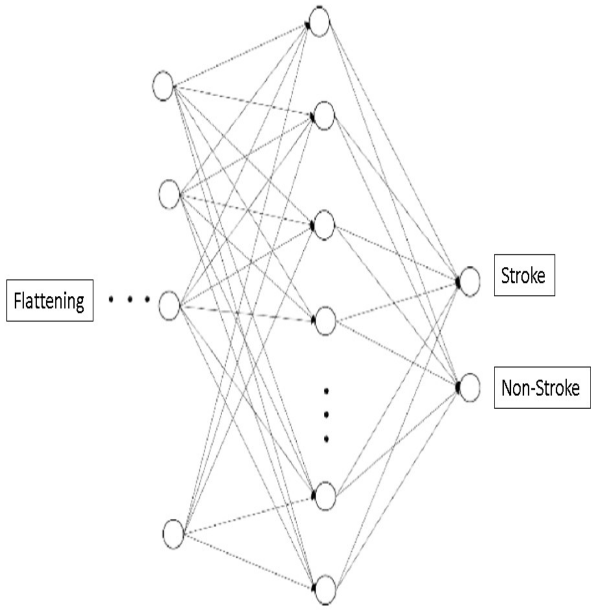

Fig. 9. Fully Connected Layer

6) Overview of the CNN-LSTM Model

An overview of CNN-LSTM Model is shown in Fig. 10. Firstly, the biosignal image dataset will be fed into the encoder to extract the features. After that, the features will be fed into the attention mechanism, and LSTM will generate refined pixels. The attention mechanism is used to assist the LSTM in learning relevant features at each step. Lastly, the refined pixels will be fed to the neural network in a fully connected layer to perform classification.

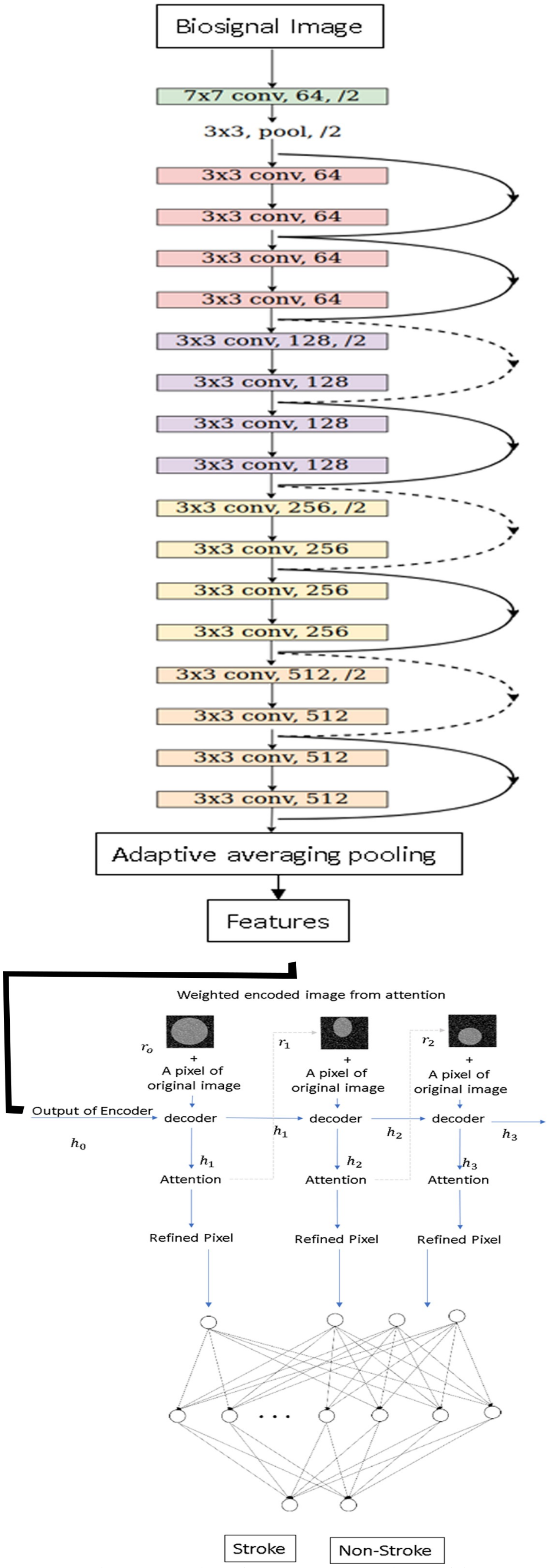

Fig. 10. Overview of CNN-LSTM Attention Model

E. Biosignal Dataset

The dataset that contained EEG, EMG and ECG signals was plotted and saved as grayscale images. The dataset was divided into a training set and a test set with a rule-of-thumb ratio of 80/20. Five-fold cross validation was applied. Fig. 11 shows the three types of biosignals, EMG, EEG and ECG in the wave form.

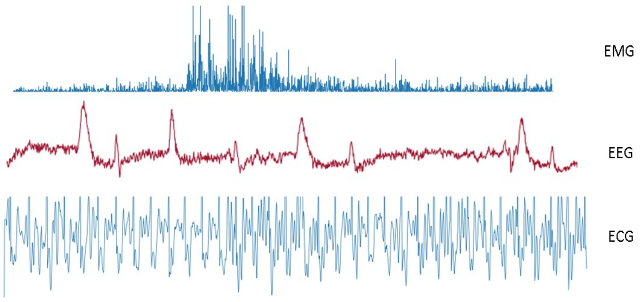

Fig. 11. Three Types of Biosignals, EMG, EEG and ECG in Wave Form

F. Interpretation of Performance Measures

When a neural network is trained, it is necessary to interpret the performance of the neural network. Therefore, accuracy, precision, recall and F1-score are used to measure the neural network's performance. The most intuitive performance indicator is accuracy. It is merely a proportion of the correctly forecast observation to the overall observations and it is a metric for evaluating classification models. Precision is the fraction of relevant instances among the retrieved instances in pattern recognition, data extraction and classification, which is shown in (18). Recall is the fraction of the total number of relevant instances that have been retrieved, which is shown in (19). F1 score can be calculated by using precision and recall shown in (20).

$$\text{Accuracy}=\frac{\text{Number of correct prediction}}{\text{Total number of prediction}}*100\% \quad (17)$$

$$\text{Precision}=\frac{\text{TP}}{\text{TP+FP}}*100\% \quad (18)$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP+FN}}*100\% \quad (19)$$

$$\text{F1 Score}=\frac{2*\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}}*100\% \quad (20)$$

G. Subject selection

Thirty people participated in this study, who were post-stroke patients and 15 healthy people. All 15 post-stroke patients fulfilled the inclusion and exclusion criteria. The inclusion criteria for stroke patients are as follows: The stroke patient must be between 30-80 years old. The cerebrovascular accident must be evidenced by radiological report or physician reports and stroke with hemiparesis with the power of the upper limb at least 2. They did not have visual problems and cognitive impairment as evidenced by the Mini-Mental State Examination score of more than 24. The exclusion criteria for stroke patients are as follows: They have cognitive impairment with a Mini-Mental State Examination score of less than 24 and the presence of uncontrolled medical illness that requires actuate medical management. They could not be enrolled in other studies targeting stroke recovery.

All 15 healthy people also fulfilled the following criteria: The age difference between healthy people and stroke patients must not be more than 10 years old and they did not have any surgery record and visual problems. Table I shows both groups of user characteristics. During the rehabilitation, they had attached EEG, EMG, and ECG devices.

All biosignal raw data were stored in a comma-separated value (CSV) format. The mini-mental examination required no additional equipment and was easy to perform. The modified Barthel index can measure physical disability by assessing the behaviour relating to activities of daily living for stroke patients or patients with other disabling conditions.

TABLE I
USER CHARACTERISTICS OF CONTROL PEOPLE AND STROKE PATIENTS

Characteristic	Control people (N=15)	Stroke patients (N=15)
Sex (male/female)	10/5	10/5
Age (yr) (mean±SD)	35±6.5	38±8.3
Dominant (left/right)	3/12	5/10
Disease duration (yr) (median)	-	3.5
First ever stroke (%)	-	90
Type of stroke (ischemic/haemorrhage)	-	11/4
Lesion side (left/right/both)	-	5/10/0
Mini mental examination (median, range)	-	25[24-26]
Modified Barthel Index (median, range)	-	52.4[44.7-56]

H. Unreal Engine (UE4)

The assessment program is developed using a game development tool called Unreal Engine 4 (UE4, version 4.24). The programming language is C++. SteamVR software is used to connect the VR software to the VR device. The UE4 coordinate system is using a 3-Dimensional Cartesian coordinate system in terms of x, y, and z [22]. The UE4 rotation system is using Euler angle. The rotation axes are in terms of roll (φ), pitch (θ), and yaw (ψ) [23]. The position of any object in UE4 can be measured in real-time. UE4 has inter-changeable codes with libraries in an object-oriented design framework. It is a computer-generated graphics system [24]. Moreover, it has a visual scripting system, called the Blueprint system [25]. The Blueprint system acts the same way as C++ classes. It has a colour code that represents the type of function. The lines in the Blueprint system are used to connect the pin of the code function. The red colour function is the header while the blue colour is the code function. The code function is represented as a node of a graph with pins for input and outputs

a) Stack the Cube (STC)

The 'Stack the Cube' game is created by using UE4 to evaluate the time taken by the user to complete the game. Fig. 12 shows the environment of STC. STC task is applied to many daily life activities. In the STC game, there are several cubes available for stacking on the table. The user is supposed to pick the cube and stack it on top of another cube. A timer starts counting after the user picks up the first cube. The timer stops counting after the player has stacked all cubes according to the requirement of the level. During the assessment, the stroke group is instructed to stack the cube with the affected hand and the control group is instructed to stack the cube with the non-dominant hand. The user should press the button of the handheld controller to pick up the cube and release the button of the handheld controller to release the cube. The dimensions of each cube are identical, 13 cm in length, width and height in a real environment. The success rate, S_{RSTC} is calculated by using (21) and expressed as percentage.

$$S_{RSTC} = \frac{n}{m} * 100\% \tag{21}$$

where, n is the number of cubes that are picked up within the 20 seconds and m is the number of cubes that are required to stack.

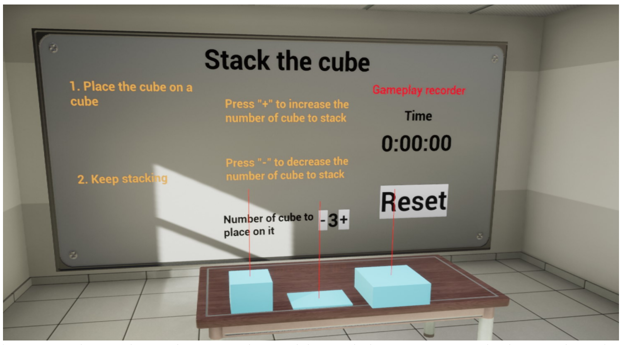

Fig. 12. The Cubes Place on the Table and the STC Board Shows the Time Taken for The Patient to Complete the Game

b) Touch the ball (TTB)

The 'Touch the ball' game is created by using UE4 and evaluates the number of balls touched by the user within 60 seconds. Fig. 13 shows the TTB gameplay. The ball has a radius of 5 cm and its speed is set at 40 cm/s. TTB is designed to assist the subject in extending the shoulder and elbow joints of the upper limb. The user must move his or her arm to touch a ball within 60 seconds. If the ball has successfully touched, the game score increases by 1. The stroke group is instructed to touch the ball with the affected hand while the control group is instructed to torch the ball with the non-dominant hand.

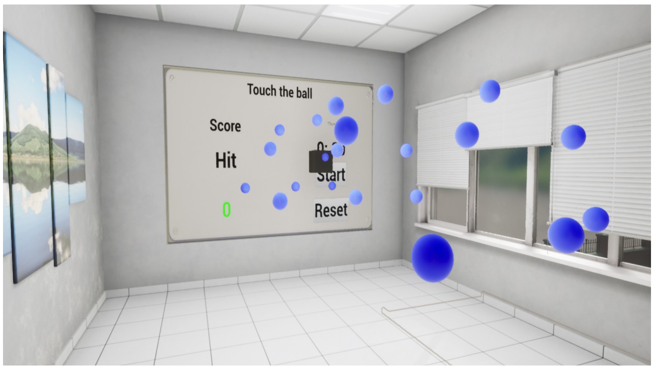

Fig. 13. The Patient Has to Touch the Blue Ball to Obtain the Score and the Dashboard Will Show the Score

The performance is evaluated based on the average score, S_A by using (22).

$$S_A = \frac{n}{5} \tag{22}$$

where, n is the number of the ball that has been touched. The average time per ball, T_{APB} is calculated by using (23) and expressed in second per ball (s/b).

$$T_{APB} = \frac{30}{n} \tag{23}$$

where n is the number of the ball that has been touched.

I. Program flow

Before starting the assessment, the biosignal devices were attached to the user's body. The user went through three phases of the biosignal measurements, i.e., pre-assessment (phase P1), during assessment (phase P2) and post-assessment (phase P3). In phase P1, the user measured the heartbeat and brain signal for 1 minute with ECG and EEG devices to obtain the resting state. During the phase P2, the user measured muscle signal and heartbeat with EMG and ECG to obtain the active state during the assessment. In phase P3, the user measured heartbeat with ECG for 1 minute to obtain the post-active state.

All games required handheld controllers and an HMD. The user sat on a chair and was familiarised with the HTC Vive. The user was given two controller devices, which he or she had to hold with their hand to interact with the virtual content. The HMD and the handheld controllers were detected by the sensors. Once the patient became familiar with the controllers, he or she started to do VR upper limb assessment. Games were played in this sequence: the 'Stack the Cubes' (STC)game followed by the 'Touch the Ball' (TTB) game. A user is required to attempt three times in the STC game and five times in the TTB game. After the user has completed all games, the biosignals at different stages of the game are extracted and displayed as images and the CNN-LSTM will be used to classify the image at each stage. The biosignals were not subjected to any pre-processing method prior to being extracted into image format. The expected duration for the user to complete each game was 25 minutes.

IV. Results

The STC game and TTB game were completed by both groups. Each person in both groups completed the games within 30 minutes. The values shown in the table are mean ± standard deviation. The error bars shown in the Fig. 14 and Fig 15 represent standard deviations

A. Stack the Cube

The mean STC time taken for the control group is 5.2±0.86 minutes as compared to 12.24±3.91 for the stroke group. Table II shows the average time taken for each level of the 'Stack the Cube' game.

TABLE II
Average Time Taken for Each Level in 'Stack the Cube' Game

Stack the cube	Control Group (seconds)	Stroke Group (seconds)
Stack 3 cubes (Level 1)	4.27±0.94	7 ±1.28

Stack 4 cubes (Level 2)	5±1.08	13.33±2.15
Stack 5 cubes (Level 3)	6.33±0.99	16.40±2.91
Overall attempts	5.2±0.86	12.24±3.91

Throughout all levels of the STC, the control group takes less time to complete all levels compared to the stroke group. Each person in both groups completed all the levels within 20 minutes. There is no difference in S_{RSTC} between the control group (100%) and the stroke patients (100%). Fig. 14 shows the average time taken by each group to complete at each level of the STC game.

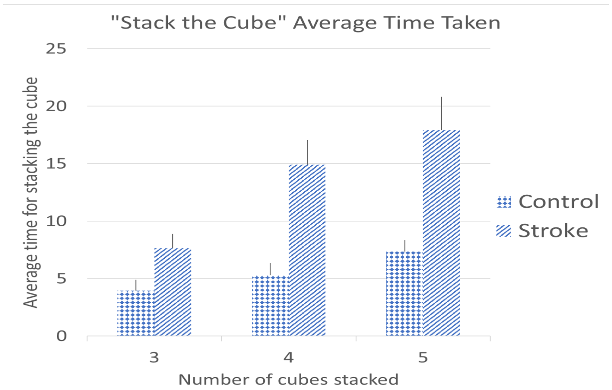

Fig. 14. Average Time Taken for Each Number of Cubes Stacked in 'Stack the Cube'

There is a significant difference in time taken between the control group and the stroke group. This shows that the stroke group takes longer time compared to the control group as they have a poor ability to move their upper limb. The time taken increases as the number of cubes that should be stacked increases. This shows that the workload for the stroke group increased as the number of cubes that were required to stack increased. The S_{RSTC} between each group is 100% because they have successfully picked and stacked the cube within 20 seconds.

B. Touch the Ball

Table III shows the average time taken for each level in TTB. Five attempts were made by all subjects from both groups. The control group has better score (20.15±0.14) than the stroke group (15.57±0.33). The stroke group had a slightly longer T_{APB} (1.93 s/b) at touching the ball than control group (1.49 s/b). Fig. 15 shows the score per attempt in "Touch the Ball" game.

TABLE III
AVERAGE TIME TAKEN FOR EACH LEVEL IN 'TOUCH THE BALL' GAME

Touch the ball	Control Group (Score)	Stroke Group (Score)
1st attempt	20.07±0.68	15.07±1.24
2nd attempt	20.27±0.44	15.73±2.14
3rd attempt	20.13±0.34	14.53±3.07
4th attempt	20.33±0.87	15.47±2.28
5th attempt	19.93±0.57	16.07±2.08
Overall attempts	20.15±0.14	15.57±0.33

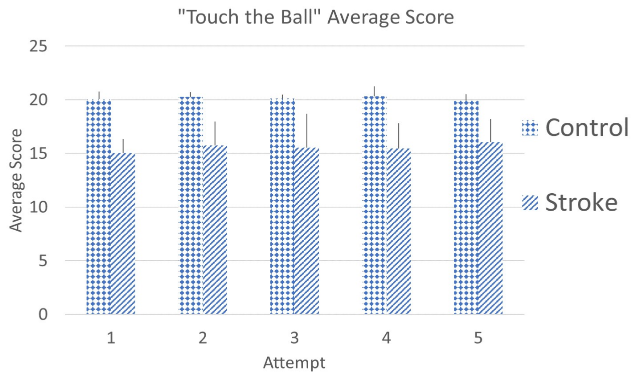

Fig. 15. Average Score for Each Attempt in 'Touch the Ball' Game.

The score of the control group is higher than the stroke group. This shows that the control group has a better ability to move the upper limb as compared to the stroke group. Moreover, due to the control group having a better upper limb motor, the control group takes less T_{APB} when compared to the stroke group. Hand movement of the control group is faster than the stroke group to touch the ball. Thus, the T_{APB} of the stroke group is longer than the control group.

C. CNN-LSTM Model Training and Validation

Table IV shows the result of the ECG classification model. Table V shows the result of the EMG classification model. Table VI shows the result of the EEG classification model. 80% of data is allocated for training and the rest is used for evaluation. The biosignals are not pre-processed and are converted into png format images. These images are collected and the condition of the participants is determined by using CNN-LSTM model. Among the three ResNet models used in this study, ResNet 50 achieved the highest accuracy compared to ResNet 34 and ResNet 18. ResNet 50 has a depth of 50 layers so it can give better results than ResNet 34 and ResNet 18. However, the ResNet 50 training time is longer compared to ResNet 34 and ResNet 18.

TABLE IV
THE RESULT ECG CLASSIFICATION COMPARED WITH RESNET MODELS

ECG ResNet	Accuracy	Precision	Recall	F1 Score
18	95	95	95	95
34	96	96	96	97
50	97	97	97	98

TABLE V
THE RESULT EMG CLASSIFICATION COMPARED WITH RESNET MODELS

EMG ResNet	Accuracy	Precision	Recall	F1 Score
18	98	98	98	98
34	97	96	97	97
50	98	98	98	98

TABLE VI
THE RESULT EEG CLASSIFICATION COMPARED WITH RESNET MODELS

EEG ResNet	Accuracy	Precision	Recall	F1 Score
18	85	85	85	85
34	86	84	83	83
50	92	90	85	82

After the comparison between ResNet 18, 34 and 50 in ECG, EMG and EEG classification, the dataset is trained with different type of models such as standard CNN, LSTM and Recurrent Neural Network (RNN) to compare with the result of the CNN-LSTM model. The first model is CNN. It is mainly used in image classification. It has three convolution operations for feature extraction and a fully connected layer for classification. The second model is LSTM. It has 3 hidden layers and 28-timesteps for classification. The third model is RNN. It has an internal memory, which can be used for memorising the previous output data and considering it with the current input. Table VII shows the result of models in ECG classification. Table VIII shows the result of models in EMG classification. Table IX shows the result of models in EEG classification. Based on Table VII, VIII and IX, the models used in the ECG, EMG and EEG classification cannot outperform the CNN-LSTM model.

TABLE VII
THE RESULT OF ECG CLASSIFICATION WITH OTHER MODELS

ECG Model	Accuracy	Precision	Recall	F1 Score
CNN	73	70	75	75
LSTM	60	61	58	60
RNN	51	54	53	55
CNN-LSTM	95	95	95	95

TABLE VIII
THE RESULT OF EMG CLASSIFICATION WITH OTHER MODELS

EMG Model	Accuracy	Precision	Recall	F1 Score
CNN	75	75	77	78
LSTM	62	60	61	65
RNN	53	51	58	54
CNN-LSTM	98	98	98	98

TABLE IX
THE RESULT OF EEG CLASSIFICATION WITH OTHER MODELS

EEG Model	Accuracy	Precision	Recall	F1 Score
CNN	71	77	79	77
LSTM	57	55	54	70
RNN	50	62	63	52
CNN-LSTM	85	85	85	85

Based on Table VII, VII and IX, CNN-LSTM has the highest accuracy, precision, recall and F1score compared to CNN, LSTM and RNN. The accuracy indicated CNN-LSTM has the highest correct prediction based on the input samples. The precision indicated quantity of CNN-LSTM predicted class is similar to the actual class. The recall indicated CNN-LSTM predicted correctly based on the number of classes in the dataset. The F1 score indicated a harmonic mean of recall and precision which CNN-LSTM has the highest score.

D. "Stack the Cube" Game Biosignal Classification

During phase P1, ECG classification classified that both groups are under slow pace condition. Fig. 16 shows the STC ECG stroke classification on both groups. Based on the EEG classification, 15 people in the control group are in the control state while 15 people in the stroke group are in the stroke state. Fig. 18 shows the STC EEG stroke classification on both groups. During phase P2, 11 people in the stroke group are under fast pace condition while 6 people in the control group are under fast pace condition. Fig. 17 shows the STC EMG stroke classification on both groups. Based on the EMG classification, 15 people in the control group are in the control state while 3 people in the stroke group are in the control state. During phase P3, ECG classification classified that 4 people in the stroke group are having fast pace condition while 10 people in the control group are having fast pace condition. Table X shows the ECG classification in STC game between stroke group and control group. Table XI shows the EMG classification in STC game between stroke group and control group. Table XII shows the EEG classification in STC game between stroke group and control group.

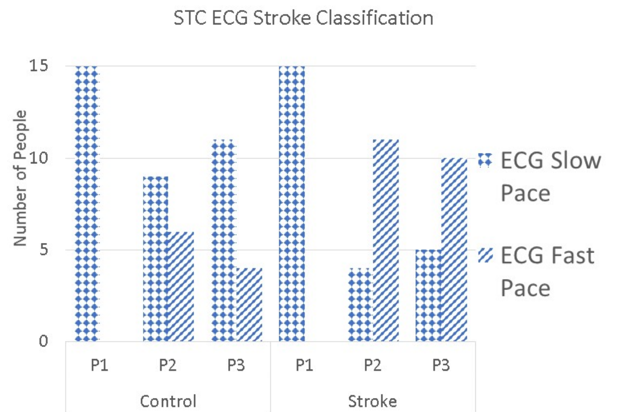

Fig. 16. ECG Stroke Classification of Stroke Group and Control Group in STC Game

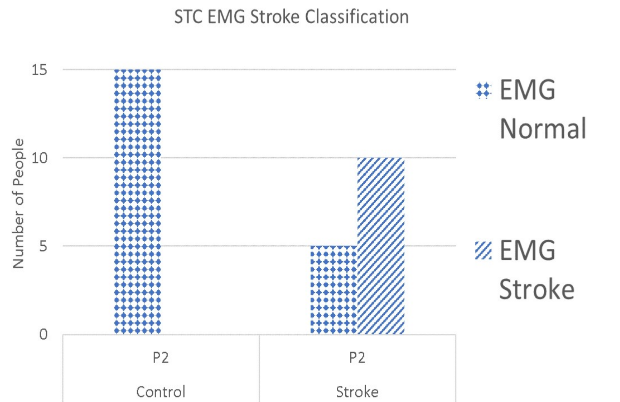

Fig. 17 EMG Stroke Classification of Stroke Group and Control Group in STC Game

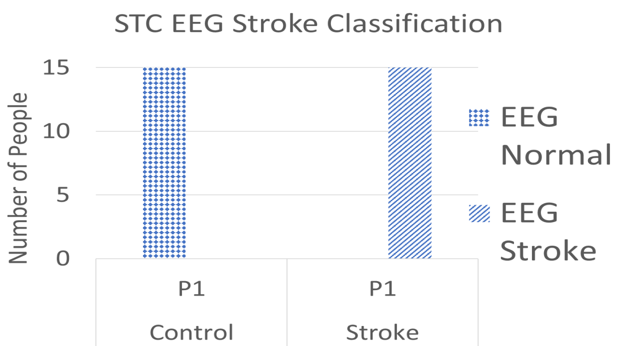

Fig. 18. EEG Stroke Classification of Stroke Group and Control Group in STC Game

TABLE X
ECG CLASSIFICATION IN STC GAME BETWEEN STROKE GROUP AND CONTROL GROUP

Group	Classification	Phase		
		P1	P2	P3
Control	Slow Pace	15	9	11
	Fast Pace	0	6	4
Stroke	Slow Pace	15	4	5
	Fast Pace	0	11	10

TABLE XI
EMG CLASSIFICATION IN STC GAME BETWEEN STROKE GROUP AND CONTROL GROUP

Group	Classification	Phase (P2)
Control	Normal	15
	Stroke	0
Stroke	Normal	5
	Stroke	10

TABLE XII
EEG CLASSIFICATION IN STC GAME BETWEEN STROKE GROUP AND CONTROL GROUP

Group	Classification	Phase (P1)
Control	Normal	15
	Stroke	0
Stroke	Normal	0
	Stroke	15

During phase P1, the neural network shows that both groups of users remained calm before performing the exercise and predicted those with and without a stroke based on EEG. During phase P2, the majority of users are calm since the STC is not a high intensity activity. Based on the EMG classification, the people in the control group are in control staste as they are able to complete STC in a short time. This means that they are able to move their upper limbs better than stroke group. In phase P3, there are less people in the stroke group are under fast pace condition compared to the people in stroke group under fast pace condition during phase P2. This means that some of the stroke patients are able to recover under slow pace condition after finished the activities.

E. "Touch the Ball" Game Biosignal Classification

During phase P1, ECG classification classified that both groups are under slow pace condition. The EEG classification states that 15 people in control group are in the control state while 15 people in stroke group are in stroke state. Fig. 21 shows the TTB EEG classification on both groups. During phase P2, ECG classification clasified that 12 people in the stroke group are under fast pace condition while 9 people in the control group are under fast pace condition. Fig. 19 shows the TTB ECG stroke classification on both groups. During phase P2, the EMG classification classified 15 people in the control group are in control state while the 6 people in the stroke group are in control state. During phase P3, ECG classification clasified that 11 people in the stroke group are under fast pace condition while 7 people in the control group are under fast pace condition. Fig. 20 shows the TTB EMG stroke classification on both groups. Table XIII shows the ECG classification in TTB game between stroke group and control group. Table XIV shows the EMG classification in TTB game between stroke group and control group. Table XV shows the EEG classification in TTB game between stroke group and control group.

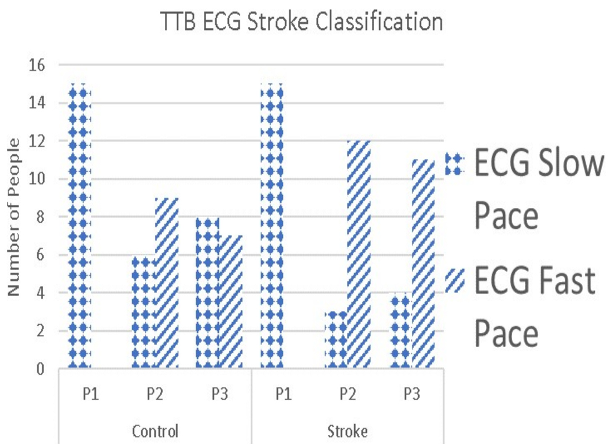

Fig. 19. ECG Stroke Classification of Stroke Group and Control Group in TTB Game

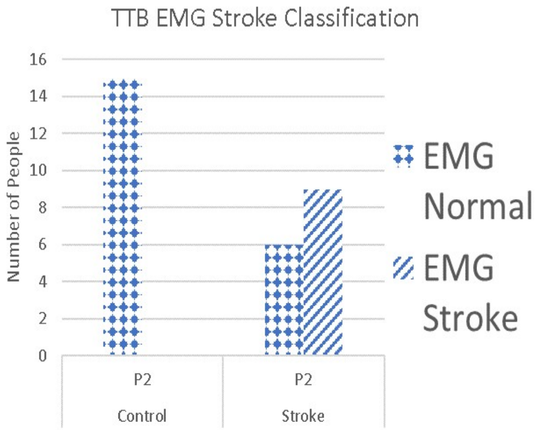

Fig. 20. EMG Stroke Classification of Stroke Group and Control Group in TTB Game

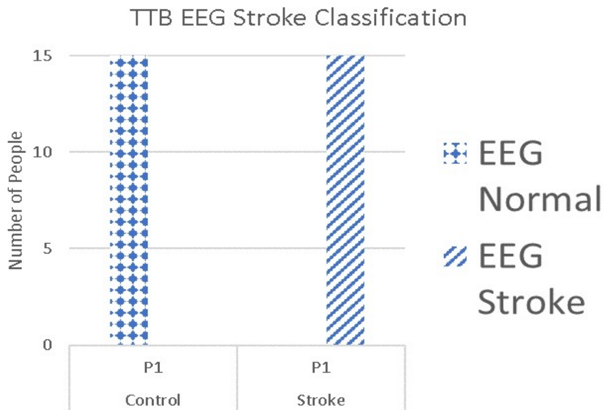

Fig. 21. EEG Stroke Classification of Stroke Group and Control Group in TTB Game

TABLE XIII
ECG CLASSIFICATION IN STC GAME BETWEEN STROKE GROUP AND CONTROL GROUP

Group	Classification	Phase		
		P1	P2	P3
Control	Slow Pace	15	6	8
	Fast Pace	0	9	7
Stroke	Slow Pace	15	3	4
	Fast Pace	0	12	11

s

TABLE XIV
EMG CLASSIFICATION IN STC GAME BETWEEN STROKE GROUP AND CONTROL GROUP

Group	Classification	Phase (P2)
Control	Normal	15
	Stroke	0
Stroke	Normal	6
	Stroke	9

TABLE XV
EEG CLASSIFICATION IN STC GAME BETWEEN STROKE GROUP AND CONTROL GROUP

Group	Classification	Phase (P1)
Control	Normal	15
	Stroke	0
Stroke	Normal	0
	Stroke	15

During phase P1 of the "Touch the Ball" game, CNN-LSTM demonstrated that both groups of users remained calm before performing the exercise and predicted those with and without stroke based on EEG image. During phase P2, most users are under fast pace condition which is understandable as TTB is a high intensity activity. Based on EMG classification in phase P2, the people in control group is in the control state as they can complete STC game in a short time. This means they have the ability to move their upper limb better than the stroke group. During the phase P3, the ECG classification classified that the stroke group has 11 people under fast pace condition. By comparing the number of people in stroke group between phase P3 and phase P2, phase P3 has less people under fast pace conditions. Thus, some of the stroke patients are able to recover under slow pace condition after doing the activities.

V. DISCUSSION

The STC game focuses on the flexion and extension movement of the upper limb as well as minor abduction and adduction movement. The flexion movement occurs when the users move their hand horizontally to pick up the cube. Abduction and adduction movement occurs when the users move their hand vertically to stack the cubes.

The focus of the TTB game is on flexion, extension, abduction and adduction movements. This is because the balls come from different positions in the same direction. Both games are designed to facilitate VR systems into the neurorehabilitation field. Due to the implementation of haptic technology in the controller, it creates a tactile experience in VR by applying force, vibration and motion to the user.

Based on the model used in this study, the stroke patient's condition is classified as a control state of either EMG, ECG or EEG, meaning that the stroke patient has a low level stroke and the stroke patient's condition is classified as the stroke state, meaning that the stroke patient has a high level stroke. Most low level stroke patients recover more quickly than high level stroke patients. This is because patients can show similar upper limb motor ability to the control group.

The ECG model, the EEG model and EMG model were trained for many times with different training settings namely learning rate, epoch size and batch size by tuning those values. The results of these models are based on the setting of learning rate at 0.001, batch size of 10 and epoch size of 40. The lower the learning rate, the more reliable the training but the optimization will take more time as the minimum of the loss function is small.

The results of ECG, EMG and EEG models with ResNet18, ResNet34 and ResNet50 are compared. The ECG, EMG, and EEG model with ResNet50 achieved the highest score while the ECG, EMG and EEG model with ResNet18 achieved the lowest score. The scores of each ResNet are relatively similar to each other. However, ResNet34 and ResNet 50 require heavy computation time and more memory compared to ResNet18. The training time for ResNet18 is shorter than ResNet34 and ResNet50. Therefore, ResNet18 has been chosen in this study.

There are several limitations to this study. Firstly, some users feel dizzy or uncomfortable prior to becoming familiar with the "Stack the Cube" game and "Touch the Ball" games. This is due to motion sickness in attempting to move around the VR platform, in addition to lack of sensor feedback when attempting to 'accelerate' in the virtual world. The way to overcome it is to give the user a break before continuing with the assessment. Secondly, the user's movement is slightly restricted due to the presence of biosignal equipment. This is because the ECG and EMG used wired sensors. Movement of the body can cause detachment of the sensors, which will create noise signals during recording. Furthermore, the signal may be distorted when the pins of the sensor move. This can be overcome by using the wireless sensor. Finally, the accuracy of the signal classification depends on the training dataset. This is important because any incorrect data being trained by the neural network will affect the accuracy of classification.

VI. CONCLUSION

In this paper, we have developed a VR upper limb assessment, namely "Stack the Cube" game and "Touch the Ball" game to evaluate performance of the user's upper limb. Moreover, biosignals devices are used during the assessment by attaching those devices on the user. The biosignals device is used to generate biosignal of the user such as heartbeat, muscle signal, and brain signal. Furthermore, a CNN-LSTM image classification has been constructed to classify the biosignal of the user either stroke or normal. The constructed model used Seq2seq modeling to learn in an end-to-end approach. We have shown the results by using different ResNet models in ECG, EMG and EEG model. Among ResNet18, ResNet34 and ResNet50, ResNet50 has the highest accuracy. However, ResNet50 used a longer time to train compared to ResNet34 and ResNet18. Thus, ResNet18 is optimum in this study. This study shows that the constructed model can effectively analyse user's biosignals and classify the biosignal either stroke or normal. Therefore, this paper can be potentially helpful in the biosignal analysis process with deep learning and the use of biosignal devices during stroke upper limb assessment. Neural network is highly accurate and precise in classifying biosignals image. The upper limb assessment can assess the user's performance based on the score obtained.

ACKNOWLEDGEMENT

The authors would like to extend their gratitude to their colleagues, Mr Fawaz Sammani and Mr Lim Choon Chen for technical support in this study and Social Security Organization (SOCSO) Rehabilitation Center Malacca for granting site visits and data collection opportunities.

REFERENCES

[1] Z. Y. Lim, K. S. Sim, and S. C. Tan, "An evaluation of left and right brain dominance using electroencephalogram signal," *Eng. Lett.*, vol. 28, no. 4, pp. 1358–1367, 2020.

[2] A. Fung, E. C. Lai, and B. C. Lee, "Usability and Validation of the Smarter Balance System: An Unsupervised Dynamic Balance Exercises System for Individuals with Parkinson's Disease," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 4, pp. 798–806, 2018, doi: 10.1109/TNSRE.2018.2808139.

[3] O. M. Giggins, U. M. C. Persson, and B. Caulfield, "Biofeedback in rehabilitation," *J. Neuroeng. Rehabil.*, vol. 10, no. 1, p. 1, 2013, doi: 10.1186/1743-0003-10-60.

[4] J. Jagannath, H. M. Saarinen, and A. L. Drozd, "Framework for automatic signal classification techniques (FACT) for software defined radios," *2015 IEEE Symp. Comput. Intell. Secur. Def. Appl. CISDA 2015 - Proc.*, pp. 54–60, 2015, doi: 10.1109/CISDA.2015.7208628.

[5] K. S. Sim, Z. Y. Lim, H. W. W. Benny, and K. T. K. Desmond, "Development of rehabilitation system using virtual reality," *Proceeding 2017 Int. Conf. Robot. Autom. Sci. ICORAS 2017*, vol. 2018-March, pp. 1–6, 2018, doi: 10.1109/ICORAS.2017.8308045.

[6] O. Marin-Pardo, A. Vourvopoulos, M. Neureither, D. Saldana, E. Jahng, and S. L. Liew, "Electromyography as a Suitable Input for Virtual Reality-Based Biofeedback in Stroke Rehabilitation," *Commun. Comput. Inf. Sci.*, vol. 1032, pp. 274–281, 2019, doi: 10.1007/978-3-030-23522-2_35.

[7] P. Trujillo *et al.*, "Quantitative EEG for predicting upper limb motor recovery in chronic stroke robot-Assisted rehabilitation,"

IEEE Trans. Neural Syst. Rehabil. Eng., vol. 27, no. 5, pp. 1058–1067, 2017, doi: 10.1109/TNSRE.2017.2678161.
[8] P. Nagabushanam, S. Thomas George, and S. Radha, "EEG signal classification using LSTM and improved neural network algorithms," *Soft Comput.*, vol. 24, no. 13, pp. 9981–10003, 2020, doi: 10.1007/s00500-019-04515-0.
[9] Z. Yan and Z. Lv, "The influence of immersive virtual reality systems on online social application," *Appl. Sci.*, vol. 10, no. 15, 2020, doi: 10.3390/app10155058.
[10] K. L. Lew, K. S. Sim, S. C. Tan, and F. S. Abas, "3D Kinematics of Upper Limb Functional Assessment Using HTC Vive in Unreal Engine 4," *Communications in Computer and Information Science*, vol. 1287. pp. 264–275, 2020, doi: 10.1007/978-3-030-63119-2_22.
[11] D. Chen, H. Liu, and Z. Ren, "Application of Wearable Device HTC VIVE in Upper Limb Rehabilitation Training," *Proc. 2018 2nd IEEE Adv. Inf. Manag. Commun. Electron. Autom. Control Conf. IMCEC 2018*, no. Imcec, pp. 1460–1464, 2018, doi: 10.1109/IMCEC.2018.8469540.
[12] V. C. Pezoulas, T. P. Exarchos, and D. I. Fotiadis, *Types and sources of medical and other related data*. 2020.
[13] J. R. Wolpaw *et al.*, "Brain–Computer Interface Technology: A Review of the First International Meeting," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 164–173, 2000, doi: 10.1109/TRE.2000.847807.
[14] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: A neural image caption generator," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3156–3164, doi: 10.1109/CVPR.2015.7298935.
[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Comput. Vis. Pattern Recognit.*, pp. 770–778, 2015, doi: 10.1002/chin.200650130.
[16] K. Xu *et al.*, "Show, attend and tell: Neural image caption generation with visual attention," *32nd Int. Conf. Mach. Learn. ICML 2015*, vol. 3, pp. 2048–2057, 2015.
[17] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.
[18] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 4, no. January, pp. 3104–3112, 2014.
[19] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
[20] K. Smagulova and A. P. James, "Overview of Long Short-Term Memory Neural Networks BT - Deep Learning Classifiers with Memristive Networks: Theory and Applications," A. P. James, Ed. Cham: Springer International Publishing, 2020, pp. 139–153.
[21] C. K. Toa, K. S. Sim, and S. C. Tan, "Electroencephalogram-Based Attention Level Classification Using Convolution Attention Memory Neural Network," *IEEE Access*, vol. 9, pp. 58870–58881, 2021, doi: 10.1109/ACCESS.2021.3072731.
[22] C. Li, A. Fahmy, and J. Sienz, "An augmented reality based human-robot interaction interface using Kalman filter sensor fusion," *Sensors (Switzerland)*, vol. 19, no. 20, 2019, doi: 10.3390/s19204586.
[23] H. A. Ardakani and T. J. Bridges, "Review of the 3-2-1 Euler Angles : a yaw – pitch – roll sequence Map from E j to a j : the yaw rotation," pp. 1–9, 2010.
[24] C. M. Torres-Ferreyros, M. A. Festini-Wendorff, and P. N. Shiguihara-Juarez, "Developing a videogame using unreal engine based on a four stages methodology," *Proc. 2016 IEEE ANDESCON, ANDESCON 2016*, 2017, doi: 10.1109/ANDESCON.2016.7836249.
[25] A. Drozina and T. Orehovacki, "Creating a Tabletop Game Prototype in Unreal Engine 4," pp. 1568–1573, 2018.