BULETINUL INSTITUTULUI POLITEHNIC DIN IAŞI
Publicat de
Universitatea Tehnică „Gheorghe Asachi” din Iaşi
Volumul 67 (71), Numărul 3, 2021
Secţia
ELECTROTEHNICĂ. ENERGETICĂ. ELECTRONICĂ
DOI:10.2478/bipie-2021-0018

HEAD GESTURE RECOGNITION BASED ON CAPACITIVE SENSORS USING DEEP LEARNING ALGORITHMS

BY

IONUȚ-CRISTIAN SEVERIN[*] and DAN-MARIUS DOBREA

“Gheorghe Asachi” Technical University of Iași,
Faculty of Electronics,Telecommunication and
Information Technology, Iași, Romania

Received: October 13, 2021
Accepted for publication: December 28, 2021

Abstract. The current paper proposed and investigated the head motion recognition idea based on four capacitive sensors and deep learning models. The proposed system was designed to empower a tetraplegic person to control a remote device or an intelligent wheelchair. The capacitive sensors were placed around the neck using a necktie, which each volunteer who participated in this experiment was easy to use. The results show that the best-proposed deep learning model can determine each activity with a classification rate equal to 89.29% using capacitive raw data. During the experiments the deep learning models provided accuracy values in the range of 56.25% to 89.29%.

Keywords: HCI system; head motion; capacitive sensors; deep learning.

[*]Corresponding author; *e-mail*: ionut.severin@etti.tuiasi.ro

1. Introduction

Human Activity Recognition and Classification are among the most exciting research fields, mainly due to the spread of wearable devices such as mobile phones and smartwatches present in our daily lives.

Determining the human body's motion and activities through wearable devices has contributed to different domains such as medicine (Severin *et al*., 2020), entertainment (Bashar, 2020), medical devices (Dobrea *et al*., 2019), elder care (Srinivasan *et al.*, 2020), and sports training (Chang *et al*., 2020). The current requirements, especially in the consumer market, are focused on achieving high classification accuracy with a low implementation cost. These requirements have become a challenging task in the last decade because most proposed wearable systems include an electronic and a sensoristic block in their structure that should satisfy the conditions of dimensionality and portability. The most recent research results available in the technical literature focus on determining the accuracy, robustness, and real-time capability. Daily activity monitoring has multiple benefits in several medical applications and has become an interesting topic for researchers worldwide (Mohmed *et al.*, 2020; Shiranthika *et al.*, 2020; Xia *et al*., 2020). The computational steps have an essential role in the final performances obtained by the proposed solutions. In the technical literature there exist two major approaches based on classical machine learning algorithms and deep learning algorithms, respectively. These categories were already studied in multiple applications from human activity recognition, focusing on the classification of general daily activity (*e.g*., reading, walking, etc.) but less in classifying specific ones (*e.g*., head motion, neck motion, etc.). In this study, the focus will be set on the head gesture classification as part of the human activity recognition domain. This information obtained from the head level is valuable in applications requiring confirmation or assistance, such as autonomous driving applications or applications that support disabled people. Therefore, our primary goal was to propose, develop, and study a new head gestural system based on lower-cost capacitive sensors. The proposed system consists of four capacitive sensors placed inside the necktie around the neck, being easy to use by each volunteer who had participated in the experiment. Another focus of the current study was to evaluate the proposed system from the perspective of classification performances provided in using deep learning algorithms. The machine-learning algorithm based on head activity recognition architectures, used in this paper, is represented by convolutional neural network-long short-term memory network (CNN-LSTM) (Shiranthika *et al.*, 2020), convolutional neural network (CNN) (Mohmed *et al.*, 2020), Long Short-Term Memory network (LSTM) (Du *et al.*, 2019), Bidirectional LSTM Network (BLSTM) (Aljarrah and Ali, 2019) and Convolutional Neural Network-Bidirectional LSTM Network (CNN-BLSTM) (Darvishzadeh *et al.*, 2019). During the current experiments all deep learning models were trained based on two new databases created by acquisition from five subjects who had neck circumference in the range of 32.5÷42 cm. The training of each deep learning algorithm was done twice over 100 and 1500 training epochs to conclude each model's classification performance. The recordings were done for each subject in two days to avoid each volunteer's dependence and obtain a result near to real life. The results show that the best-proposed deep learning model can determine each activity with a classification equal to 89.29% using capacitive raw data. Also, during the experiment, the deep learning models provided accuracy in the range of 56.25% to 89.29%. This study provides a good overview of the less expensive non-contact wearable sensor classification capability that can have an excellent potential to be included in diverse human-computer interaction (HCI) solutions. The current study's main objective was to design, investigate, and evaluate the classification performance provided by classic and hybrid deep learning models capable of recognizing each predefined head motion command with accuracy. Also, this study provides a good starting point for further research, where such sensors can be included in intelligent systems to facilitate the interaction between humans and smart devices as an alternative to complex or expensive sensors.

2. Related Work

Human motion analysis generated a significant interest from worldwide researchers in the last decade. After inspecting the relevant literature on the subject we concluded that few applications monitor and analyze specific parts of the human body. The possible reason for the fewer studies focused on the specific parts of the human body is related to the availability of proper, wearable sensor types. In this paper, the obtained results were compared with our previous results based on inertial sensors (Severin and Dobrea, 2020), which the international community already validated. If the head motion classification is considered part of the human activities classification, many studies propose and evaluate different methods to classify it in the technical literature. In (Bashar *et al.*, 2020), human activity recognition (HAR) is investigated based on a smartphone and neural network model. Their study was oriented to classify six daily activities. The proposed model is evaluated on a public data set and achieved 95.79% classification accuracy. Another experiment performed in (Mohmed *et al.*, 2020) proposes a fuzzy feature representation approach to Convolutional Neural Network for human activity modeling and recognition. During their experiment, the body motions were gathered from the environmental sensors (*e.g.*, passive infrared sensor, pressure sensor, temperature sensor, etc.). The overall classification accuracy obtained a value equal to 97.8%. In both cases, the obtained performances are excellent. Still, the principal disadvantage is the reduced portability or impossibility of detecting a specific motion performed by a single body part. In the technical literature, many handcrafted feature extractions are explored to improve the classification performance of computational models. For this reason, it becomes complicated to get high performances with the increasing diversity of the used wearable sensors. Moreover, multiple topologies have been proposed and studied. The automatic feature extraction and classification can represent an advantage provided by the deep learning model approach. Most existing methods incorporate a deep network that attempts to extract features from the acquired time series. Shiranthika *et al.*, 2020 studied the possibility of using the CNN and LSTM architectures for human activity recognition problems. Based on the published results, the CNN model was capable of getting 99.53% training accuracy and 99.46% validation accuracy. In comparison, the LSTM model got a classification accuracy equal to 84.71%. These results were obtained based on the WISDM dataset, known in the literature as smartphone and smartwatch activity and biometrics datasets. This dataset has data acquired from 51 test subjects who performed 18 activities for 3 min at a 20 Hz sampling rate. Another study conducted by (Yen *et al.*, 2020) studied the possibility of recognizing six daily human activities (walking, walking upstairs, walking downstairs, sitting, standing, and lying) using a deep learning algorithm. The proposed systems were based on a single inertial sensor containing a three-axis accelerometer and a three-axis gyroscope. The computational steps consist of a feature-learning method based on a 1D convolutional neural network that automatically performs feature extraction and classification from raw data. After those computational steps were accomplished, they obtained a training classification accuracy equal to 98.93% and 95.99% testing accuracy. Besides classical deep learning architectures, recently, several new hybrid architectures had appeared. Such architecture was treated in (Xia *et al.*, 2020). They propose a hybrid neural network that combines a long short-term memory network (LSTM) with a convolutional neural network (CNN). In their experiment, the computational models were fed into a two-layer LSTM followed by convolutional layers. Additionally, a global average pooling layer was applied to replace the fully connected layer. To train the model, three public databases were used. The reported results suggest that the overall accuracy of their model achieves a value equal to 95.85%. The results published in the field of human activity recognition are motivating based on deep learning techniques. In our previous studies, we treat the head motion problem based on inertial sensors (Severin and Dobrea, 2020). Our main goal was to propose, design, and implement the state-of-the-art head gestural system that has a good balance between cost and wearability in our earlier studies. The proposal studies have multiple implications and can be beneficial in the medical field providing support for disabled persons. The obtained results based on the inertial sensor are promising and are used as reference results for current studies where a new type of non-contact sensor is used.

A new system used for head gestural detection based on lower-cost capacitive sensors was evaluated in this study. The proposed system is a new real-time, low-power, and non-contact wearable system designed to determine the head motion having the possibility to be included in multiple human-computer interaction interfaces (*e.g.*, an interface for controlling an intelligent wheelchair). The capacitive sensors were placed symmetrically around the neck disposed against the sagittal plane. The proposed system was used during the experiment by each volunteer involved in the measurements from two days. The obtained dataset was used to train the deep learning predictive models. The focus was to evaluate and determine the best deep learning models that can accurately recognize the head motion based on the capacitive time series. The deep learning models used in this experiment are LSTM, BLSTM, CNN, CNN-BLSTM, and CNN LSTM. These five deep learning algorithms were already designed and used in one of our previous study (Severin and Dobrea, 2020), where each of them was trained based on inertial data set. The obtained results based on the inertial sensor highlighted that the classification accuracy of the best inertial deep learning models is equal to 93.64%. Also, each predefined head gestural command was classified in the range of 71.11% to 100%. These results were used to validate the proposed system and determine the advantages and disadvantages provided by the capacitive sensors. The data acquisition was performed on two days during the experiment with a sampling frequency rate equal to 10 Hz. In the current investigation, six predefined commands were used to be classified by each deep learning model. Each of these predefined commands had recorded from five volunteers who were instructed how to perform each head motion correctly. Each of them executes the head motion sitting on the chair with the backrest without moving the body from the neck down. The final databases obtained after two days of the experiment contain 28051 data sampling for the first day and 22200 for the second day. The offline analyses were performed in PyCharm 2017.1.4 Professional version. The training model tasks were performed on a Lenovo IdeaPad 700 laptop with an Intel Core i7 processor (CPU @2.59GHz, 4-cores, 8-logical Processors) and 24 GB physical memory. The main contribution of this paper is given by the: proposal, analysis, and evaluation of the head gesture classification model based on a new approach that uses a lower cost capacitive sensor. The second contribution brought by this experiment is given by the performance analysis of the five deep learning models adapted to working with capacitive time series. This approach provides a good overview of the potential of using the lower-cost capacitive sensors in the head gesture classification. The final contribution provided by the current studies is related to the inspection of the possibility to determine the head motion independently as part of human activity with a lower cost sensor type. This approach is beneficial in medical areas or facilitates the interaction between humans and computers (HCI). Also, the current paper treats the condition of inspection of a single part of the human body instead of monitoring the motion performed by the entire human body, which is a common approach in technical literature.

3. Experimental Setup

The current study uses the capacitive time series provided by four capacitive sensors placed inside a necktie. The proposed wearable system (Dobrea *et al.*, 2019) contains capacitive sensors, an FDC2214 driver circuit, and a SimpleLink CC2650 wireless board used to receive the capacitive signals associated with the head motion. The FDC2214 driver has the primary role of converting the analogical values from all four capacitive sensors to numerical values. This circuit is a 28-bits capacitance-to-digital converter. The SimpleLink CC2650 wireless board contains three main component blocks. These are represented by a CPU block, an RF core, and a sensor controller engine. A diagram block representation of the used system for head motion detection can be seen in Fig. 1 (Dobrea *et al.*, 2018):

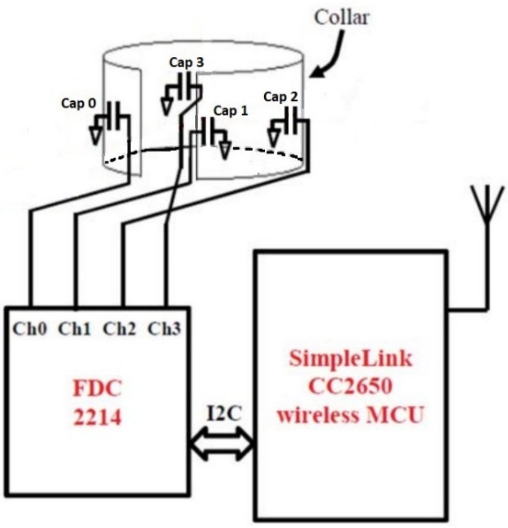

Fig. 1 – Block diagram of Head Gestural motion system.

After the acquisition, the capacitive time series are sent using Bluetooth protocol to the processing window application developed in previous research. The databases used in this experiment to train the deep learning models were created under the acquisition step from five volunteers with a neck circumference placed in a range of 32.5÷42 cm. The recordings of capacitive time series were done for each volunteer on two different days, except volunteer five, for which only on the first day of experiment data was acquired. The head motion was categorized into six predefined head gestural commands. Each predefined head motion command was performed by each volunteer sitting on the chair without moving the body from the neck down, simulating a spastic tetraparesis patient. The predefined commands are represented by motion forward, backward, left, right, stop and a random one interpreted as no-command type. Each predefined head motion commands were performed below (Dobrea *et al.*, 2019):

- "*forward*" command (A1): the inclination of the head to the left side, followed by an inclination to the downward;
- *"backward"* command (A2): an inclination of the head to the left side, followed by an inclination to the upwards;
- "*left*" order (A3): an inclination of the head to the left side, followed by the rotation of the head to the left;
- "*right*" command (A4): a tilt of the head to the right side, followed by a rotation of the head to the right;
- "*stop*" command (A5): an inclination of the head to the upwards, then, to the downward, followed by the rotation of the head to the left and right;
- "*garbage*" command (A6): recordings of the head motion for 3.5 seconds, with a random behaviour and no-command type.

Each volunteer who had part of this experiment performed 30 recordings for each of the six predefined head motions at a sampling frequency equal to 10 Hz. During the investigation, all the processing and analysis of the head classification performance was performed offline. The training of the deep learning models was done based on the raw form of the capacitive time series without any other preprocessing steps. The final form of the used gestural system can be observed in Fig. 2 (Dobrea *et al.*, 2019):

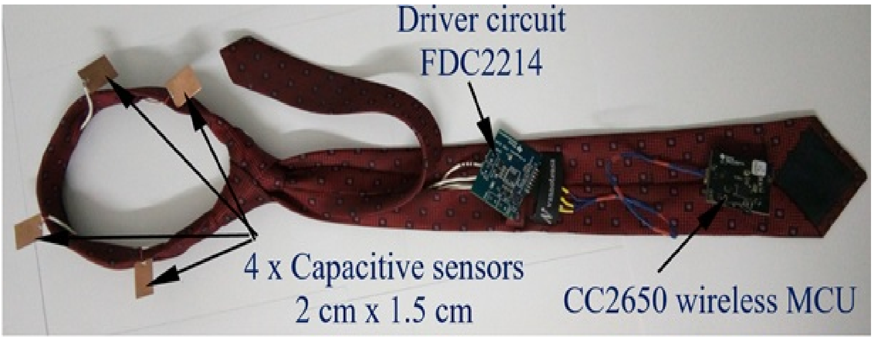

Fig. 2 – Head Gestural motion system based on capacitive sensors.

4. Deep Learning Models

For this experiment, five neural network algorithms were used to determine the Head Gesture classification accuracy rate. The utilized algorithms are represented by LSTM, BLSTM, CNN, CNN-BLSTM, and CNN BLSTM. These five algorithms were chosen in such a way to work well with the time series, which in this paper are represented by the value of capacitive sensors.

Convolutional neural network (CNN)

This neural network algorithm is one of the most used and efficient deep learning algorithms in the application based on the human activity task classification. The CNN architecture involves several layers of processing, using multiple elements in parallel operation. During this experiment, to classify the head gesture more accurately, it was adopted to be used tools such as pooling, rectified linear unit (ReLU), activation function, and soft-max classifier usually used in several deep learning tasks. Such a deep learning algorithm performs the convolution step instead of matrix multiplication. The main advantages provided by this deep learning method are related to the fact that the model offers the possibility to be trained using raw time series values without any preprocessing steps. Based on this deep learning approach, the head gesture features are extracted automatically through-alternating use of the convolutional layer and pooling layer from the original capacitive time series. For this experiment, the CNN neural network was implemented in Python using the Keras library. The input for the neural network was reshaped to be interpreted as a 1D image. To assure this, each capacitive time series was sampled in fixed-width sliding windows of 8.6 sec and 50% overlap (86 readings/window). As an activation function for this deep learning model, it was used the rectifier activation function. Also, for this experiment, the CNN model was specified to have a single fully connected layer. The size of the used CNN neural network can be seen in Fig. 3.

Layer (type)	Output Shape	Param #
conv1d_1 (Conv1D)	(None, 85, 128)	1152
max_pooling1d_1 (MaxPooling1	(None, 42, 128)	0
dropout_1 (Dropout)	(None, 42, 128)	0
conv1d_2 (Conv1D)	(None, 42, 128)	32896
max_pooling1d_2 (MaxPooling1	(None, 21, 128)	0
dropout_2 (Dropout)	(None, 21, 128)	0
flatten_1 (Flatten)	(None, 2688)	0
dense_1 (Dense)	(None, 6)	16134

Total params: 50,182
Trainable params: 50,182
Non-trainable params: 0

Fig. 3 – Head Gestural motion CNN model.

The pooling layer has a primary role in the proposed model to reduce the feature map dimensions and the total number of nods. The Dropout layer has the role of preventing overfitting and accelerating the training process by removing invaluable nodes. The Dense layer represents the last used layer. This layer has the main function of adding the connected layers.

Long Short-Term Memory (LSTM)

Another Neural Network algorithm used in this paper is represented by LSTM, an extension of the Recurrent Neural Network. This algorithm is working well for the case when a time series is used for the classification side. This one avoids the vanishing and exploding gradient problems. The LSTM architecture contains two inputs that determine the direction of data flow from left to right. Those two inputs are represented by the current input x_t and the previous cell output h_{t-1}. An LSTM network also contains many cell blocks that have different components. The input gate, forget gate, and output gate represent these components. The architecture of such a neural network can be seen in Fig. 4:

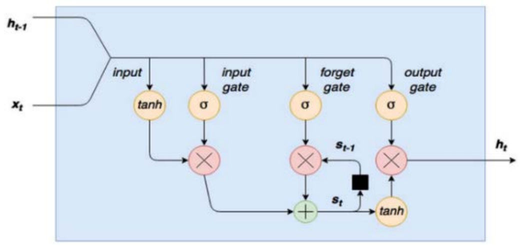

Fig. 4 – LSTM architecture (Kusumo, 2018).

For this experiment, the LSTM neural network was implemented in Python using the Keras API library. The used neural network (NN) was configured to have 128 hidden layers and two fully connected layers. First, fully connected layers contain 256 nodes, while the second one has 128 nodes. As an activation function for the fully connected layer was used the rectifier activation function. For the output layer, the activation function used was SoftMax. For the training, part was used 80%, while 20% was used for the testing model.

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 128)	68096
dropout_1 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 256)	33024
dense_2 (Dense)	(None, 128)	32896
dense_3 (Dense)	(None, 6)	774

Total params: 134,790
Trainable params: 134,790
Non-trainable params: 0

Fig. 5 – Head Gestural motion LSTM model.

Fig. 5 is highlighted the used LSTM Head gestural model implemented with Keras API.

Bidirectional LSTM Network (BLSTM)

This neural network algorithm is a new approach based on the LSTM model. This approach proposes that each training sequence forward and behind are two LSTM cells. This thing leads to the fact that the new structure can calculate each cell's past and future states in the input sequence. In the new architecture, the hidden layer saves two values. One component from the hidden layer is used for forwarding calculation, while the second is used for reverse calculation (Yang, 2019). In this experiment, the Bidirectional LSTM network was configured to have 128 hidden layers. For the output part, similar to the LSTM Neural Network were used SoftMax activation function. The architecture of this type of neural network used during this experiment can be seen in Fig. 6:

```
Layer (type)                 Output Shape              Param #
=================================================================
bidirectional_1 (Bidirection (None, 86, 256)           136192
_________________________________________________________________
dropout_1 (Dropout)          (None, 86, 256)           0
_________________________________________________________________
flatten_1 (Flatten)          (None, 22016)             0
_________________________________________________________________
dense_1 (Dense)              (None, 6)                 132102
=================================================================
Total params: 268,294
Trainable params: 268,294
Non-trainable params: 0
```

Fig. 6 – Head Gestural motion BLSTM model.

Convolutional Long Short-Term Memory network (CNN-LSTM)

The fourth used neural network algorithm for this experiment is represented by a hybrid one called CNN-LSTM. This type of neural network contains two main components: CNN, and the second one is represented by LSTM. Such a neural network was born because it was observed that CNN has good performance for extracting the primary information from the input data, while the LSTM was much better than CNN to predict the input sequences. Based on CNN, the feature of input data is automatically extracted with the help of alternation between the 1D convolutional layer and the pooling layer. In the convolutional layer part, each convolutional kernel is specialized to extract a specific part of the time series feature. The pooling layer has the primary role of selecting the most important k features characteristics determined at the convolution layer step. In the model used during this research, the fully connected layer was not used. During the experiment, this model was configured to have two convolutional layers with 128 filters. The kernel size was configured to be two, while the activation function is rectifier function. The kernel size was chosen to have a small computation cost and the best classification performance. For our subsequent research, we have in-plane to determine the optimal value of the kernel size based on the trial experiments. The convolutional block was designed to have two pooling layers with a window size equal to 2. After each pooling layer, a dropout layer (20%) was included to perform the regularization step and ovoid overfitting data. For this experiment, the batch size was configured to be equal to 5. The architecture of this type of neural network used during this experiment can be seen in Fig. 7:

Layer (type)	Output Shape	Param #
conv1d_1 (Conv1D)	(None, 85, 128)	1152
max_pooling1d_1 (MaxPooling1	(None, 42, 128)	0
dropout_1 (Dropout)	(None, 42, 128)	0
conv1d_2 (Conv1D)	(None, 42, 128)	32896
max_pooling1d_2 (MaxPooling1	(None, 21, 128)	0
dropout_2 (Dropout)	(None, 21, 128)	0
lstm_1 (LSTM)	(None, 128)	131584
dropout_3 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 6)	774

Total params: 166,406
Trainable params: 166,406
Non-trainable params: 0

Fig. 7 – Head Gestural motion CNN-LSTM model.

Convolutional Neural Network - Bidirectional LSTM Network (CNN - BLSTM)

The fifth used neural network algorithm is represented by a hybrid one, which is called CNN-BLSTM. This type of neural network contains two main components: the CNN model and the second one represented by the BLSTM model. The CNN component has the same role in the final architecture as was specified in the previous section; when using alternation between the 1D convolutional layer and the pooling layer, the features of the input time series were automatically extracted. This architecture combines the structure of the convolutional layer with the bidirectional LSTM model considering the time series of a past event and a future event. The structure of the model was configured in the current experiment, similar to the structure of the CNN-LSTM model described in the previous chapter. The data input for the proposed model was sampled in fixed-width sliding windows of 8.6 seconds (86 readings/window). In the current experiment, each capacitive time series was normalized before training the deep learning models. The used mathematical relation for the standardization step is:

$$IC\_new = \frac{IC\_old - \mu}{\sigma} \tag{1}$$

where:

IC_new: capacitive time series new

IC_old: initial capacitive time series

μ: mean value

σ: standard deviation value

The architecture used during this experiment for the estimation of the head gesture activity can be seen in Fig. 8:

Layer (type)	Output Shape	Param #
conv1d_1 (Conv1D)	(None, 85, 128)	1152
max_pooling1d_1 (MaxPooling1	(None, 42, 128)	0
dropout_1 (Dropout)	(None, 42, 128)	0
conv1d_2 (Conv1D)	(None, 42, 128)	32896
max_pooling1d_2 (MaxPooling1	(None, 21, 128)	0
dropout_2 (Dropout)	(None, 21, 128)	0
bidirectional_1 (Bidirection	(None, 21, 256)	263168
dropout_3 (Dropout)	(None, 21, 256)	0
flatten_1 (Flatten)	(None, 5376)	0
dense_1 (Dense)	(None, 6)	32262

Total params: 329,478
Trainable params: 329,478
Non-trainable params: 0

Fig. 8 – Head Gestural motion CNN-BLSTM model.

According to the presented models, the total number of parameters for each proposed head gestural deep learning (HGDL) model were 134.790 (LSTM), 166.406 (CNN-LSTM), 329.478 (CNN-BLSM), 268.294 (BLSTM), and 50.182 (CNN). Each HGDL model was implemented in Python 3.7 and Keras deep learning library version 2.3.1. In this study, compared to most already published studies, the focus was to determine the head gestures based on interpreting the capacitive time series. This approach is new in head gestural detection and can be an alternative to conventional ones (video camera, radar sensor, etc.).

5. Experimental Results and Discussion

The evaluation metrics used at offline evaluation are represented by the confusion matrix, training accuracy, test accuracy, train loss, and test loss. The obtained results based on the capacitive sensors presented during this paper were compared with the results obtained in another already presented paper (Severin and Dobrea, 2020). The head motion was interpreted based on the inertial sensor. Each HGDL model was trained with two distinct databases obtained based on data acquisition from two different days. For the training part was used 80%, while 20% was used for the testing part. After 100 and 1500 training epochs, the best training accuracy was obtained using the hybrid models represented by the CNN-LSTM and CNN-BLSTM. For both cases, the models can learn each predefined command with an accuracy of up to 94.16% over 100 training epochs on databases (DB1) obtained in the first experimental session. For the second case, when the predictive models were trained based on the databases obtained on the second day of the experimental session (DB2), we obtained an accuracy equal to 100% at the final of 100 training epochs for both models. The classification accuracy obtained on the testing DB1 over 100 epochs was equal to 72.99% for CNN-BLSTM and 85.40% for CNN-LSTM. With the testing DB2, the classification accuracy was equal to 83.93% (CNN-BLSTM) and 89.29% (CNN-LSTM). Besides the previously mentioned deep learning models, another two models have good classification accuracy, higher than 72% (DB1) and 78% (DB2). This one is represented by BLSTM (72.26%), and CNN (75.18%) applied on DB1. While applied on DB2, the BLSTM model has an accuracy equal to 81.25%, and CNN accuracy is equal to 78.57%.

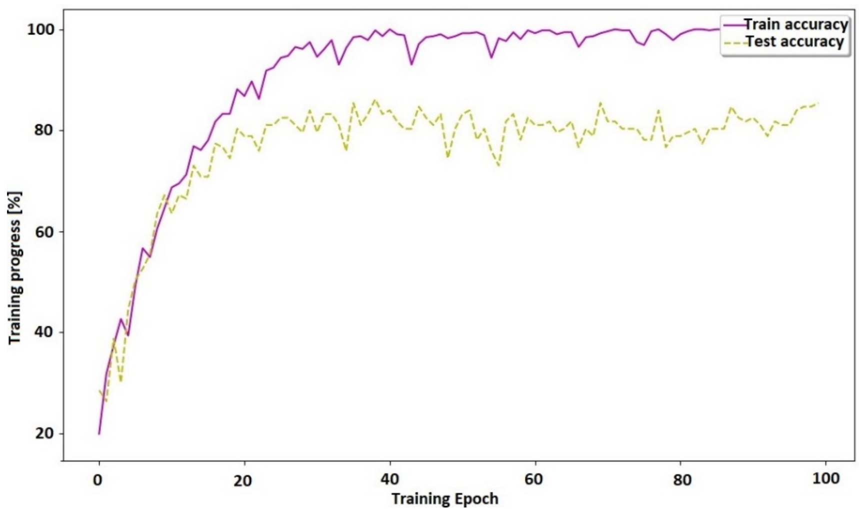

Fig. 9 – Classification accuracy over 100 epochs for CNN-LSTM with DB1.

The lowest classification accuracy during the experiment was obtained with the LSTM neural network and had the value equal to 65.69% (DB1) and 56.25% (DB2). In Fig. 9 and Fig. 10 is presented the evolution of training and testing classification performance over 100 epochs for the CNN-LSTM model:

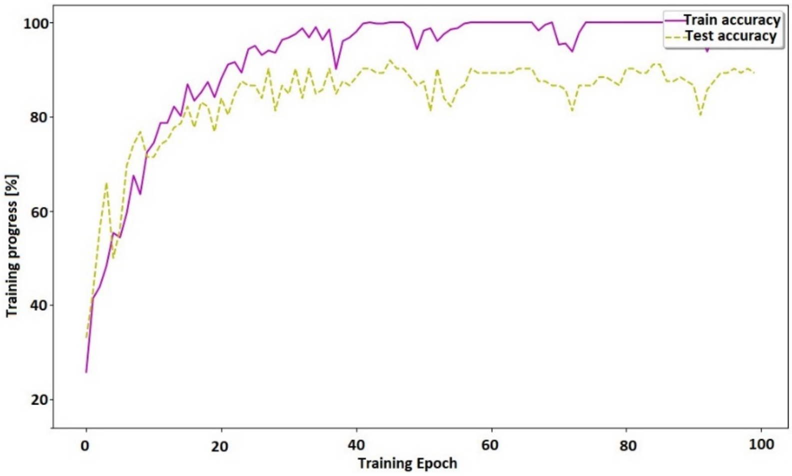

Fig. 10 – Classification accuracy over 100 epochs for CNN-LSTM with DB2.

In the case of 1500 training epochs, the deep learning models confirmed the previously presented results. This suggests that the proposed models are stable and are not affected by the neck circumference of each user. The obtained results for the best deep learning model case can be seen in Fig. 11 and Fig. 12.

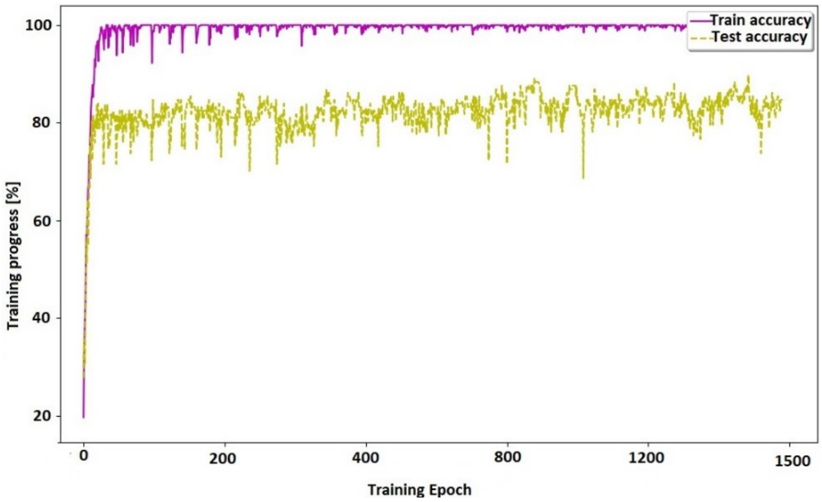

Fig. 11 – Classification accuracy over 1500 epochs for CNN-LSTM with DB1.

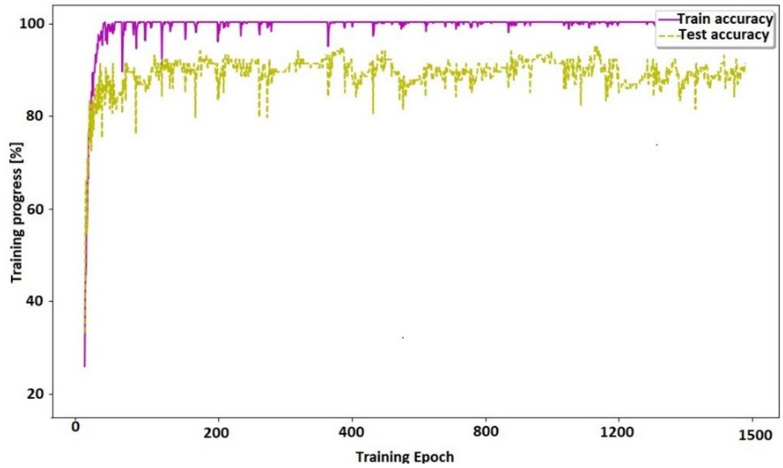

Fig. 12 – Classification accuracy over 1500 epochs for CNN-LSTM with DB2.

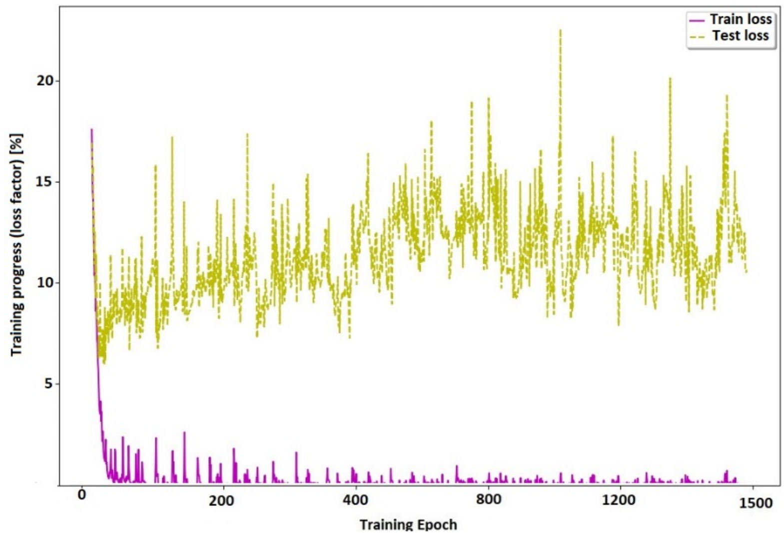

Fig. 13 – Loss factor over 1500 training epochs for CNN-LSTM with DB1.

After 1500 training epochs, the classification accuracy for the rest of the predictive models is equal to 79.56% (DB1) and 81.25% (DB2) for LSTM, 72.99% (DB1), and 79.46% (DB2) for CNN, while for CNN-BLSTM is 77.37% (DB1) and 85.71% (DB2). Based on the previously presented results, it was concluded that the deep learning models are stable. Also, this capability is confirmed by the evolution of the loss factor over the 1500 training epochs. In Fig. 13 is observed that the loss factor for the test set is in the range of 6% to 17% for the worst case. In the case of the training set, the loss factor is around 4% for the first 200 epochs and decreases to 1-2% for the rest of the training epochs. The results provided by the loss factor suggest that the proposed head gestural system can be capable of correctly detecting the head motion during 8-repetition and 2 wrong from a total of 10-repetition. In this experiment, the evaluation of the capability of each deep learning model to classify the predefined commands was made with the help of Confusion Matrix. This tool provides the possibility to monitor the classification rate for each head gestural predefined motion. In Fig. 14 and Fig. 15 is evidenced the confusion matrix obtained for the best deep learning model represented by CNN-LSTM.

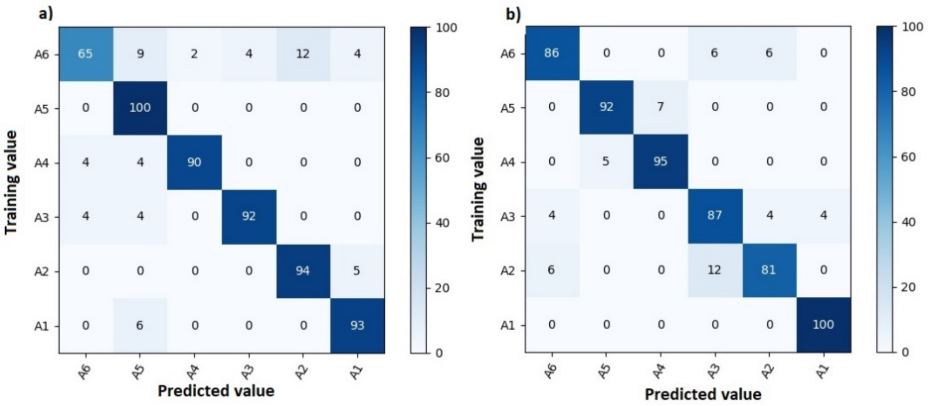

Fig. 14 – Confusion matrix obtained over 100 training epochs for CNN-LSTM with: a) DB1 and b) DB2.

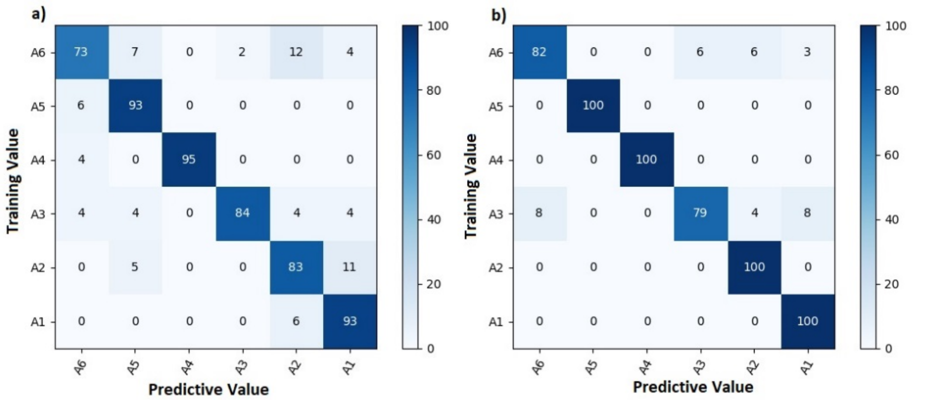

Fig. 15 – Confusion matrix obtained over 1500 training epochs for CNN-LSTM with: a) DB1 and b) DB2.

Based on the Confusion matrix presented in Fig. 14 and Fig. 15, it is suggested that the deep learning model is capable of identifying each six predefine head motions with an accuracy of up to 65% (over 100 epochs) and 73% (over 1500 epochs). The less classification performance was obtained in the case of identification of random head motion (no-command). Based on the results presented, the proposed head gestural system suggests that it can be a promising alternative for the classical ones (*e.g.*, video camera, ultrasonic sensor, etc.). To validate the current system, we compare this one with a previous proposal system based on a single inertial sensor (Severin and Dobrea, 2020). The comparative results are presented in Table 1 for each deep learning model used in the current study and a previous already published study (Severin and Dobrea, 2020).

Table 1

Results were obtained on the 20% test set and 80% training set for the capacitive sensor and inertial sensor over 100 epochs

Classifier	Capacitive Training [%]	Capacitive Testing [%]	Inertial Training [%]	Inertial Testing [%]
LSTM	78.21	65.69	98.83	88.18
CNN	99.81	75.18	97.03	82.12
CNN-LSTM	100	85.40	99.30	92.42
BLSTM	100	72.26	100	91.12
CNN-BLSTM	94.16	72.99	100	93.64

In Table 1, for the case of capacitive system is used DB1, to compare the performance with the inertial system (Severin and Dobrea, 2020). Based on Table 1, we observed that in the case of best deep learning models (CNN-LSTM or CNN-BLSTM) the classification performance in the case of the capacitive sensor system is less with 7% (CNN-LSTM) and 20% (CNN-BLSTM) than inertial system studied in a previously published study (Severin and Dobrea, 2020). Although even the system studied in the current paper obtains a classification rate lower than the system proposed in another study, this one provides several advantages. Few of them are represented by low development cost, easy inclusion in wearable clothes, and good classification rate. The performances obtained for each deep learning model for both head gestural systems (capacitive vs. inertial sensor) can be observed in Fig. 16.

Comparing actual results with the previous one, when the current head gestural system was evaluated with classical machine learning algorithms (Dobrea, 2019), it is suggested that the best classification is 7% higher than in the best classification deep learning model. The good results, obtained in (Dobrea, 2019), were favored by the additional preprocessing step, while the current study was excluded (each model was trained with raw time series).

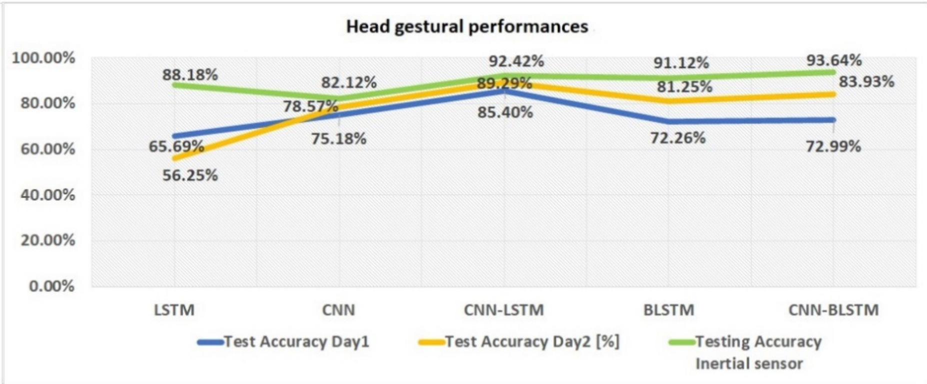

Fig. 16 – The classification performances over 5 deep learning models.

For further research, we consider extending the current models using an additional preprocessing step to improve the classification rate. From a previous study (Dobrea, 2019) and the current one was observed that capacitive time series could be used in recognition of head motion with a good classification rate from both perspectives, classical machine learning vs. deep learning models.

6. Conclusion

During this study, it was proposed and investigated a new low-cost gestural system. The computational steps were performed using 5 deep learning models that were trained to recognize six head gestural motions applied to persons who suffer from tetraplegia or quadriplegia. After two training sessions (100 and 1500 epochs), it was managed to achieve a maximum test accuracy equal to 89.29% (CNN-LSTM). Besides this predictive model, the other three got an accuracy higher than 75%. These are represented by CNN (78.57%), BLSTM (81.25%), and CNN-BLSTM (83.93%). The lowest results have been received with LSTM predictive model. For this model, the results suggest that they can recognize each predefined head gestural motion with accuracy in a range of 56.25% to 65.69%. These results were obtained over training sessions of each predictive model with two databases acquired over two experimental days from 5 volunteers. For the best predictive model, the test error factor had a value equal to 10.71%. This experiment evaluated the possibility of integrating deep learning models to recognize head gestural motion using capacitive time series as an alternative to the classical approach (*e.g.*, video camera, radar sensor, etc.). Based on the obtained results, the proposed system is promising to develop new human-computer interaction systems. For further study, we intend to improve the proposal system from the SW and HW points of view. From the SW viewpoint, we intend to add a preprocessing step based on dynamic time warping (DTW) to cope with different head motion speeds. Another intention is to extend the size of databases using two approaches. The first approach involves the acquisition of data from multiple volunteers, while the second one is based on using the GAN model to generate new training or test data. Another interesting task we plan to do is analyze the possibility of head motion segmentation from the capacitive time series containing multiple behaviors simultaneously (*e.g.*, lower and upper limb motion, chest motion, etc.). Based on this scenario, we will determine and provide a good overview of the possibility of integrating this head gestural system into more intelligent and complex systems that will require learning a new set of head gestures automatically.

REFERENCES

Aljarrah A., Ali A., *Human Activity Recognition Using PCA and BiLSTM Recurrent Neural Networks*, Proc. Internat. Conf. on Engineering Tech. and its Appl., IICETA 2019, 156-160.

Bashar S., Al Fahim A., Chon K., *Smartphone Based Human Activity Recognition with Feature Selection and Dense Neural Network*, Proc. Annual Internat. Conf. of IEEE EMBC Society, 2020, 5888-5891.

Chang W., Hsu C., Chen L., Su J., Chen M., *A Wearable Devices-Based Home Sports Recording System for Health Management*, Proc. IEEE Internat. Conf. on Consumer Electronics, ICCE-Taiwan 2020, 1-2.

Darvishzadeh A.*et al*., *CNN-BLSTM-CRF Network for Semantic Labeling of Students' Online Handwritten Assignments*, Proc. Internat. Conf. on Document Analysis and Recognition, ICDAR 2019, 1035-1040.

Dobrea M., Dobrea D., Severin I., *A New Wearable System for Head Gesture Recognition Designed to Control an Intelligent Wheelchair*, Proc. E-Health and Bioeng. Conf., EHB 2019, 1-5.

Dobrea D., Dobrea M., *A Neuronal Model of the 3D Head Position Based on a Wearable System*, Proc. Internat. Conf. and Exposition on Electrical and Power Eng., EPE 2018, 0341-0346.

Du Y., Lim Y., Tan Y., *Activity Prediction using LSTM in Smart Home*, Proc. IEEE 8-th Global Conf. on Consumer Electronics, GCCE 2019, 918-919.

Kusumo B., Heryana A., Nugraheni E., Rozie A., Setiadi B., *Recognizing Human Activities and Earthquake Vibration from Smartphone Accelerometers Using LSTM Algorithm*, Proc. Internat. Conf. on Computer, Control, Informatics and its Applications, IC3INA 2018, Tangerang, Indonesia, 88-92.

Mohmed G., Lotfi A., Pourabdollah A., *Convolutional Neural Network Classifier with Fuzzy Feature Representation for Human Activity Modelling*, Proc. IEEE Internat. Conf. on Fuzzy Systems, FUZZ-IEEE 2020, 1-7.

Shiranthika C., Premakumara N., Chiu H.-L., Samani H., Shyalika C., Yang C.-Y., *Human Activity Recognition Using CNN & LSTM*, Proc. 5-th Internat. Conf. on Information Tech. Research, ICITR 2020, 1-6.

Severin I., Dobrea D., *Head Gesture Recognition Based on 6DOF Inertial Sensor Using Artificial Neural Network*, Proc. Internat. Symp. on Electronics and Telecomm., ISETC 2020, 1-4.

Severin I., Dobrea D., Dobrea M., *Head Gesture Recognition using a 6DOF Inertial IMU*, Internat. J. of Computers Comm. & Control, **15**, (2020).

Srinivasan A. *et al*., *Elder Care System Using IoT and Machine Learning in AWS Cloud*, Proc. IEEE 17-th Internat. Conf. on Smart Communities: Improving Quality of Life Using ICT, IoT and AI, HONET 2020, 92-98.

Xia K., Huang J., Wang H., *LSTM-CNN Architecture for Human Activity Recognition*, IEEE Access, **8**, 56855-56866 (2020).

Yang S., Research on Network Behavior Anomaly Analysis Based on Bidirectional LSTM, Proc. IEEE 3-rd Inf. Tech., Networking, Electronic and Automation Control Conference, ITNEC 2019, Chengdu, China, 798-802.

RECUNOAŞTEREA GESTURILOR DE LA NIVELUL CAPULUI UTILIZÂND SENZORII CAPACITIVI ȘI ALGORITMII CU ÎNVĂȚARE PROFUNDĂ

(Rezumat)

În lucrarea actuală s-a propus și investigat ideea de recunoaștere a mișcării capului prin intermediul utilizării a patru senzori capacitivi și a algoritmilor cu învățare profundă. Sistemul propus a fost conceput astfel încât să poată oferi posibilitatea unei persoane tetraplegice să controleze un dispozitiv media sau un scaun cu rotile inteligent de la distanță. Senzorii capacitivi au fost plasați în jurul gâtului, folosind o cravată, fiind ușor de utilizat de fiecare voluntar care a participat la acest experiment. Rezultatele arată că cel mai bun model cu învățare profundă (CNN-LSTM), poate determina fiecare activitate predefinită cu o precizie egală cu 89,29%. De asemenea, pe parcursul experimentului, toate modelele cu învățare profundă au avut o precizie cuprinsă între 56,25% și 89,29%.