IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 30, 2019, accepted September 17, 2019, date of publication September 30, 2019, date of current version October 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2944652

A Novel Combination Model of Convolutional Neural Network and Long Short-Term Memory Network for Upper Limb Evaluation Using Kinect-Based System

LEI CAO [1,2], (Member, IEEE), CHUNJIANG FAN [3], HAORAN WANG [2], AND GUOKAI ZHANG [2]
[1] Department of Computer Science and Technology, School of Information and Engineering, Shanghai Maritime University, Shanghai 201306, China
[2] Department of Computer Science and Technology, Tongji University, Shanghai 201804, China
[3] Wuxi Rehabilitation Hospital, Wuxi 214043, China

Corresponding authors: Chunjiang Fan (951786665@qq.com) and Guokai Zhang (lcao@shmtu.edu.cn)

This work was supported by the MS201758, Dai Kerong Academician Workstation.

ABSTRACT Nowadays, somatosensory devices were widely used for developing the rehabilitation systems for limb-injured patients. However, professional evaluation was rarely studied in this field. In our study, we presented a novel hybrid deep network combined long short-term memory (LSTM) network and convolutional neural network (CNN) for rehabilitation evaluation referred to Brunnstrom Scale. In the identification task of 3-class Brunnstrom stages (III, IV, V), the mean accuracy of our proposed model was up to 80% (84.1%). The experimental result validated the reliability of our proposed evaluation method. And the comparison result of three machine learning algorithms indicated that the superiority of our hybrid model for Kinect-based 3D data analysis.

INDEX TERMS Rehabilitation evaluation, hybrid deep network, LSTM, CNN, Brunnstrom Scale.

I. INTRODUCTION

In many countries, spinal cord injury (SCI) is an extremely pervasive health-care problem, causing death and acquired physical disability [1]. SCIs lead to partial or full paralysis and result in lasting loss of motor function because damaged axons can not regenerate and the death of a considerable quantity neurons occur in the injured spinal cord [2], [3]. People with such behavioral deficit experiences dramatic pain in performing daily activities such as dressing, washing, and eating. Relevant neuroplasticity studies have indicated that SCI patients with physical deficiencies can be partially recovered through proper physiotherapy rehabilitation [4], [5]. However, when patients perform therapy exercises at home, they can not maximize the therapeutic benefits of activity-based therapy due to lack of supervision and evaluation. Besides, usually, rehabilitation assessment of BCI patients adopts this combination of filling in scales and professional judgment of therapists [6]–[8]. The evaluation methods are heavily depending upon the visual assessment of physical therapists, according to international standards (e.g., The International Standards for the Neurological Classification of Spinal Cord Injury (ISNCSCI)). Meanwhile, the performance may be not accurate enough for several reasons, one of which is the subjectivity of these clinical observations and evaluation. Moreover, large amounts of money are paid to Physical Therapist (PT) or Occupational Therapist (OCT), which increases the financial burden to families of patients.

It is against this background that new rehabilitation approaches based on Virtual Reality (VR) are being developed and have been widely applied in the field of augmented BCI rehabilitation in recent years because of the reduction of associated labor, providing repeatable rehabilitation movement training, and accurate evaluation of rehabilitation performance [9]. The critical technology of virtual reality-based rehabilitation (VRBR) is to use sensor tools to capture and quantify the movements of patients for monitoring their progress precisely [10]. Burdea et al. (1997) were the first to

The associate editor coordinating the review of this manuscript and approving it for publication was Linbo Qing.

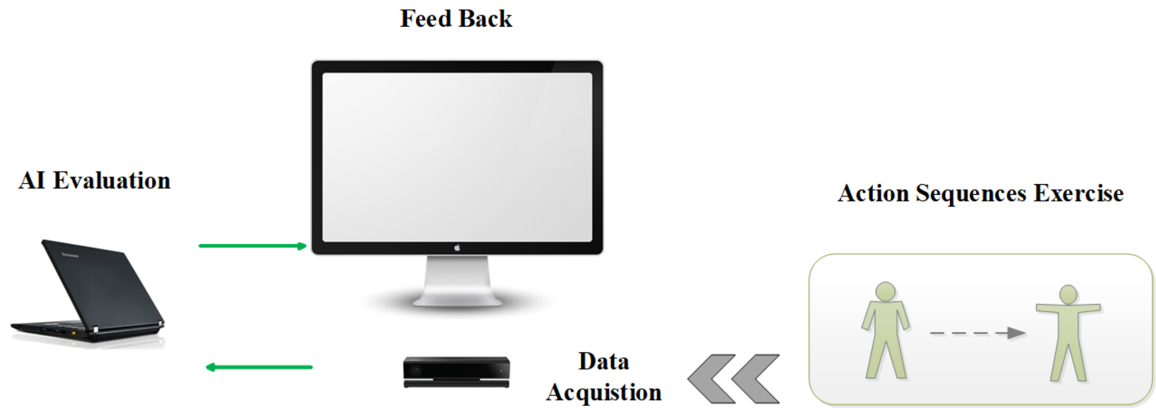

FIGURE 1. The system paradigm of the Kinect-based rehabilitation evaluation software. The flow chart of the Kinect-based evaluation process was given in this figure.

integrate VR technology into hand diagnosis and rehabilitation [11]. During performing exercises modeled on the basis of standard hand rehabilitation, patient movement state was fed back using the Rutgers Master worn on the patient's hand in the experiment. And after that, VRBR was incorporate into or was used to replace conventional rehabilitation. Some studies add extra VRBR devices to obtain outcomes better than traditional standard rehabilitation in many respects such as promoting balance [12]–[14], improving walking ability [15]–[17] and mobility [18], [19].

As the critical component within VRBR, motion capture (MoCap) via Microsoft Kinect sensor currently has become increasingly popular in physical therapy and rehabilitation, because Kinect does not require users to wear any marker compared with other sensors [20], [21]. The Kinect sensor captures depth images as well as color images at a high frame rate [22]. Color images taken by RGB camera are used in facial and body recognition. The sensor was used for measuring the distance of each joint point of the skeleton by transmitting near-infrared light and calculating the distance of bounced light propagation [23], [24].

Lots of researches have proved the feasibility of applying the Kinect-based Virtual Reality system to upper limb rehabilitations. Chang et al. rehabilitated two adolescents with cerebral palsy using a Kinect-based system. Results indicated that the two participants significantly increased their upper-limb mobility and improved exercise performance [25]. Hueso et al. introduced a customized Kinect-based VR system which allows users to implement cognition and physical rehabilitation therapies. Therapists are able to configure the treatment remotely and send them to a personal computer that belonged to patients. The system is used for collecting motion information while patients performed rehabilitation exercises [20].

Nowadays, artificial intelligence (AI) is used in nearly every medical fields for improving the effectiveness of rehabilitation software [26]. For example, Weng et al. (2017) applied four training algorithms (support vector machine (SVM), gradient boosting methods, neural networks) on clinical samples collected from more than 370,000 patients free from cardiovascular disease to predict first cardiovascular event over 10-years. Models were evaluated by statistical analysis (e.g., Area Under Curve (AUC), sensitivity, negative predictive value (NPV)). The best-performing algorithm was Neural network which has higher prediction accuracy compared to the established algorithm (American College of Cardiology guidelines) [27]. Rajkomar et al. constructed a deep learning algorithm composed of three neural network model architectures (LSTM, an attention-based model, and a neural network with boosted time-based decision stumps). Then it was validated the novel method using electronic health record data from two US academic medical centers [28].

In this paper, we proposed a novel rehabilitation evaluation system (RES) based on the Kinect sensor and employed deep-learning networks combined LSTM with CNN models to assess the patients. Specifically, with the aid of Kinect-based skeleton tracking, while patients performed exercises, body movement was recognized, and the 3D coordinates data were recorded for further analysis. Furthermore, we trained the hybrid model on 3D data to classify rehabilitation phases of patients.

II. MATERIALS AND METHODS

A. SYSTEM OVERVIEW

The system is composed of a desktop computer, a 3D Kinect camera and a 40-inch display (Fig. 1). In the evaluation process, it presents a video where an instructor to perform correct actions designated by a professional therapist according to Brunnstrom Scale: (1) Touching Homolateral Ear; (2) Shoulder Extension 90 degrees; (3) Shoulder Extension 180 degrees; (4) Scapular Retraction; (5) Shoulder Adduction; (6) Shoulder Flexion 90 degrees; (7) Shoulder Flexion 180 degrees; (8) Hand Rotation; and (9) Shoulder Back Extension (see TABLE 1 for detailed explanations).

The rehabilitation system shows patients a series of standard movements on the screen in the using period. The patient imitates the instructor's actions on the screen to perform the corresponding exercise. In this way, every patient completes

TABLE 1. **The detailed description of 9 action sequences designed by PT according to BRUNNSTROM Scale assessment.**

ID	Action	Description
Action 1	Touching Homolateral Ear	Shoulder raise, flexion and adduction from arm above
Action 2	Shoulder Extension 90 degrees	Arm outstretched in front of body with shoulder at 90 degrees flexion, elbow extended to arm outstretched at side at 90 degrees abduction, elbows extended
Action 3	Shoulder Extension 180 degrees	Arm outstretched in front of body with shoulder at 180 degrees flexion, elbow extended to arm outstretched at side at 180 degrees abduction, elbows extended
Action 4	Scapular Retraction	Arms extended in front of body at 45 degrees flexion, elbows extended - flexing elbows to bring hands back towards body
Action 5	Shoulder Adduction	Shoulder flexion and adduction from arm above head across midline of body towards opposite knee with elbow extended
Action 6	Shoulder Flexion 90 degrees	Shoulder flexion 90 degrees with elbow extended
Action 7	Shoulder Flexion 180 degrees	Shoulder flexion 180 degrees with elbow extended
Action 8	Hand Rotation	Rotation of the hand with elbow at 90 degrees flexion
Action 9	Shoulder Back Extension	Arm outstretched in back of body with shoulder at 45 degrees flexion

all nine tasks described in Table 1. These tasks are designed by a professional therapist with the purpose of rehabilitating upper-limb mobility and frequently employed in SCI rehabilitation and diagnostics. Each task is repeated twice by each participant. Meanwhile, the skeletal data captured by the system via Kinect sensor to represent patient's movements are stored into the dataset. The result of the second time was used for results evaluation.

B. ALGORITHM ARCHITECTURE

Our classification methods is a hybrid deep network combined by two conventional algorithms, LSTM network and CNN. The overall neural network architecture which we construct for classification of rehabilitation is shown in Fig. 2. It composes primarily of two different neural network architectures, LSTM and CNN. Moreover, we use two other neural networks, Support Vector Machine (SVM) and Random Forest (RF) for comparison of classification performance. All classifiers perform 3-class classification tasks in our experiment.

1) LONG SHORT TERM MEMORY (LSTM)

LSTM is a noticeable variation of Recurrent Neural Network (RNN) that was designed to overcome the shortcoming of conventional RNNs in the long-term dependence problem [29]. The memory cell c_t, which makes up LSTM neural networks and can memory and forget context information, is the major innovation in contrast to simple RNNs. The LSTM cell updates its state information by controlling the activation of the gates. Formally, the mathematic formulas of the update of an LSTM unit are as follows,

$$i_t = \sigma_i(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (1)$$

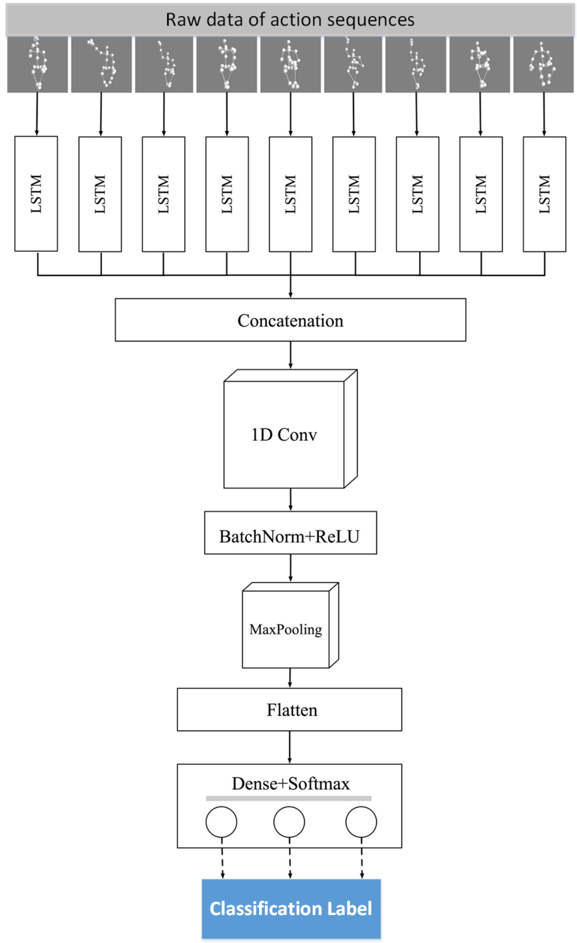

FIGURE 2. **The system paradigm of our Kinect-based rehabilitation evaluation.**

$$f_t = \sigma_f(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (2)$$

$$c_t = f_t c_{t-1} + i_t\sigma_c(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma_o(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (4)$$

$$h_t = o_t\sigma_h(c_t) \quad (5)$$

where i, f, o, and c are respectively the input gate, forget gate, output gate and cell activation vectors, all of which have the same size as the hidden state vector h (also acts as output) storing all the useful information, and σ is the element-wise activation function [30]. The term W is the weight matrix of different gates, with subscripts which indicate from-to relationships (e.g., W_{xi} being weight matrix for input x and W_{hi} being weight matrix for hidden-input h), and b represents bias vector [31].

2) CONVOLUTIONAL NEURAL NETWORK (CNN)

Convolutional neural networks have made tremendous progress in the field of image recognition [32]. CNN can be regarded as a particular type of back-propagation network neural. Compared to the standard neural network, CNN introduces convolution and pooling operations to achieve shift and deformation invariance. Its design concepts are mainly based on three ideas: local receptive fields, shared weights, and spatial subsampling. Local receptive fields mean that a CNN cell receives inputs from a subset of previous layer's cells located in a small neighborhood [33]. CNN runs a small size window (also so-called convolution kernel) that is convolved with a local region data over the input so that weights of the network can extract features from the data. This way in which one convolutional layer extracts the same feature using the corresponding window at every position of input data is referred to as shared weights [34]. Once an abstract feature has been extracted, the global one becomes unimportance, as long as the relative relationship can be preserved. Hence every convolutional layer is commonly followed by pooling layer which carries out averaging and subsampling operations, reducing feature dimension, avoiding overfit and favorable for classification.

3) BATCH NORMALIZATION

Batch normalization (BN) is the most widely adopted approach that improves the training speed of Deep neural networks (DNNs) and stabilizes the distribution of inputs by reducing internal covariate shift. Internal covariate shift is a phenomenon that the parameter changes of the previous layers cause the distribution changes of each layer inputs in the training process. It makes training Deep neural networks complicated.

Let an m size mini-batch $B = \{x_{1...m}\}$ be input, $\hat{x}_{1...m}$ be the normalized values, and $y_{1...m}$ be the output of linear transformations. The algorithm of BN is described as follows:

- Mean of feature in mini-batch:

$$\mu_B \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \tag{6}$$

- Variance of feature in mini-batch:

$$\sigma_B^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_B)^2 \tag{7}$$

- Normalize feature of samples:

$$\hat{x}_i \leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \tag{8}$$

- Linear scale and shift:

$$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv BN(x_i) \tag{9}$$

where ϵ is a hyper parameter to set in advance. On the contrary, the parameters γ and β are learned for optimizing. The normalized values $\hat{x}$ has the expected value of 0 and the variance of 1, and the scaled and shifted value y is passed to the next layer.

4) HYBRID NETWORK OF LSTM AND CNN

The proposed neural network architecture for assessment of patient rehabilitation phases is shown in Fig. 2. The whole network is initialized with random parameters. The input of the model is three-dimensional data of nine actions in cartesian coordinates. The learning rate is 5e-4. The data of one action is taken into a 25-unit LSTM in the first layer of the network. The facilities are conducive to obtain the sequences related information as optimal features. The outputs of each LSTM are concatenated and then are passed onto a 1D convolutional layer, followed by Batch Normalization which is used to prevent the overfitting problem from occurring and makes training more stable for the parameter scale. After the Batch Normalization, ReLU activation is used to bring nonlinear components and also make the model have high expressive power [35], [36]. The convolution operation routinely followed by a max-pooling strategy, which applies downsampling on each filter for subsample to reduce the output dimensionality [37]. Finally, the output of the max-pooling layer was flattened into a 1D vector, and the vector is delivered to a softmax classifier through dense connections.

C. COMPARISON ALGORITHMS

1) SUPPORT VECTOR MACHINE

SVM is a common machine learning technique initially designed to solve classification problems. Given the training set $S = \{(x_1, y_1), \ldots, (x_l, y_l)\}$, where $x_i \in R^d$ and $y_i \in \{-1, +1\}$. The SVM model maps the vectors x_i into higher dimensional space by the kernel function $\varnothing$ and attempts to find an optimal separating hyperplane $W^T\varnothing(x_i)+b = 0$ with the maximal margin to distinguish two kinds of samples. The SVM model can be formulated as follows [38]:

$$\begin{aligned} &\min_{w,b,\xi} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{l}\xi_i \\ &s.t.\ y_i(\boldsymbol{w}^T\varnothing(\boldsymbol{x_i}) + b) \geq 1 - \xi_i \\ &\quad \xi_i \geq 0. \end{aligned} \tag{10}$$

$C > 0$ is the penalty factor of the error term, which can trade off training error and balance model complexity. In this paper, we used the radial basis function (RBF) is used as the kernel function and its formula is presented as following,

$$K(x_i, x) = exp(-\frac{|x_i - x|^2}{2\sigma^2}) \tag{11}$$

Here, σ^2 is the the kernel parameter.

2) RANDOM FOREST CLASSIFIER

Random forest classifier is an ensemble classification method that the multiple decision trees are integrated. It uses bagging to generate these decision trees, in which a random sampling of the training dataset is applied to each tree [39]. Hence predictions of these trees have low correlation thereby avoid over-fitting. In addition, during the training phase, Random forest classifier randomly selects a subset of features to grow a decision tree. Then these trees predictions are aggregated to provide a final prediction.

D. SUBJECTS AND EXPERIMENTAL SETTINGS

In our study, 23 subjects performed (14 males and 9 female, aged from 43 to 56) were recruited by the doctors at Wuxi

TABLE 2. The comparable classification results for three models. A 3-fold cross validation were used for evaluating the experimental performance.

Method	CV1 (%)	CV2 (%)	CV3 (%)	Mean Value (%)
SVM	57.1	33.3	66.7	55.7
RF	42.9	66.7	66.7	58.8
Hybrid model	85.7	83.3	83.3	84.1

Rehabilitation Hospital. After confirming the approval of informed consent, they participated in further evaluation tasks by Kinect-based system. Neither of the participants had previous experience with Kinect-based system. At last, 19 valid data were recorded finally. Also, two veteran PTs were invited to evaluate the Brunnstrom stage for each patient.

The acquired data of one subject contained 9 trials of action sequences listed as Table 1. The subject needed to complete the prescribed action in 20 seconds for a trial. In this one, $26 \times 3 \times 250$ joint points were collected by Kinect sensor for further analysis. The doctors would assess the 3-class Brunnstrom stages (III IV V) for all patients after the experiment. The numbers of three classes were 6, 6 and 7 respectively. 3-fold cross validation (CV) method was used for evaluating the accuracy of three different models.

III. RESULTS

In our study, three classification models were used for evaluating rehabilitation stages. Moreover, two indicators were used for evaluating the classification results. (1) Accuracy rate (ACC): the percentage of successful selections of Brunnstrom stages assessed by PTs. (2) Receiver operating characteristic (ROC) curves: a plot of test sensitivity as the y coordinate versus its 1-specificity or false positive rate as the x coordinate.

A. CLASSIFICATION ACCURACY

Table 2 showed the comparison results of three classification methods. Average CV ACCs indicated that the precisions of all algorithms were higher than the random level of 3-class classification (i.e., 33.3%). Moreover, our proposed hybrid model was most excellent in the performance of Brunnstrom stages assessment.

B. ROC CURVES

In Fig. 3, we listed the ROC curves of every class for three models. Generally, the area under the curve of the hybrid model was larger than those of other classifiers for each category. Especially for Brunnstrom stage III, the true positive rate of hybrid model reached the ideal value of 1. These findings varied that the hybrid model was feasible for Kinect-based movement data analysis in the field of rehabilitation engineering.

IV. DISCUSSION

In our study, we proposed a novel combination model of LSTM and CNN for rehabilitation assessment. The Kinect-based camera was used for data acquisition. The experimental result demonstrated that our hybrid model was significantly better than other classification methods. The probable reason was correlated with the structure of deep neural network. CNN was suitable for extracting local features in the 3D space. And LSTM was able to solve time series tasks. Hence, the features of non-stationary 3D coordinate series could be effectively extracted by the hybrid models.

Another advantage of the hybrid model was its tolerability against missing data. In our experiment, four participants (2 persons at Brunnstrom stage III, two persons at Brunnstrom stage IV) had not finished all action tasks caused by extreme impairment of the upper limb. About forty percents of motion data could not be collected for further analysis. They were assigned to 0 for data preprocessing. Finally, the ROC curves suggest that the hybrid model had higher fault-tolerant capability than those of other classifiers. It was demonstrated that the robustness of our proposed model was reliable.

Previously, a lot of studies focused on the development of relevant Kinect-based exercise and game application [40], [41]. It was proved that these technologies were reliable for movement rehabilitation. However, few works presented somatosensory interactive -based systems for professional rehabilitation evaluation [42], [43]. Though the accuracy of their experimental result was higher than 90%. They didn't provide reasonable pieces of evidence corresponding to the professional measure paradigm. And our designed evaluation system was developed by Brunnstrom Scale. It was objective and credible for users. Furthermore, we firstly performed multi-class classification for practical application. It was useful for the patients under treatment.

The distribution of 3 classes was balanced. While the experimental results showed that the under area of ROC curves represented Brunnstrom stage V was larger than those of other categories for each classification models. It was implied that this category was easier to distinguish among them. The result was in line with our expectations. The patients diagnosed at the Brunnstrom stage V could perform the daily movement of the upper limb without any assistance. The self-care abilities of them were superior to those of the disables diagnosed at the Brunnstrom stage IV and III. For instance, Fig. 4 illustrated the action trajectories of Action 2 from 3-class Brunnstrom stages. Subject 3 took less time to complete the single motion task. And the line of the third patient was more regular than those of other subjects. It was revealed that Brunnstrom stage V was easier to distinguish from path tracking by Kinect-based sensors.

As introduced, our evaluation system was feasible for precise measurement without any assistance by professional doctors. It is confirmed that this device would be exploited

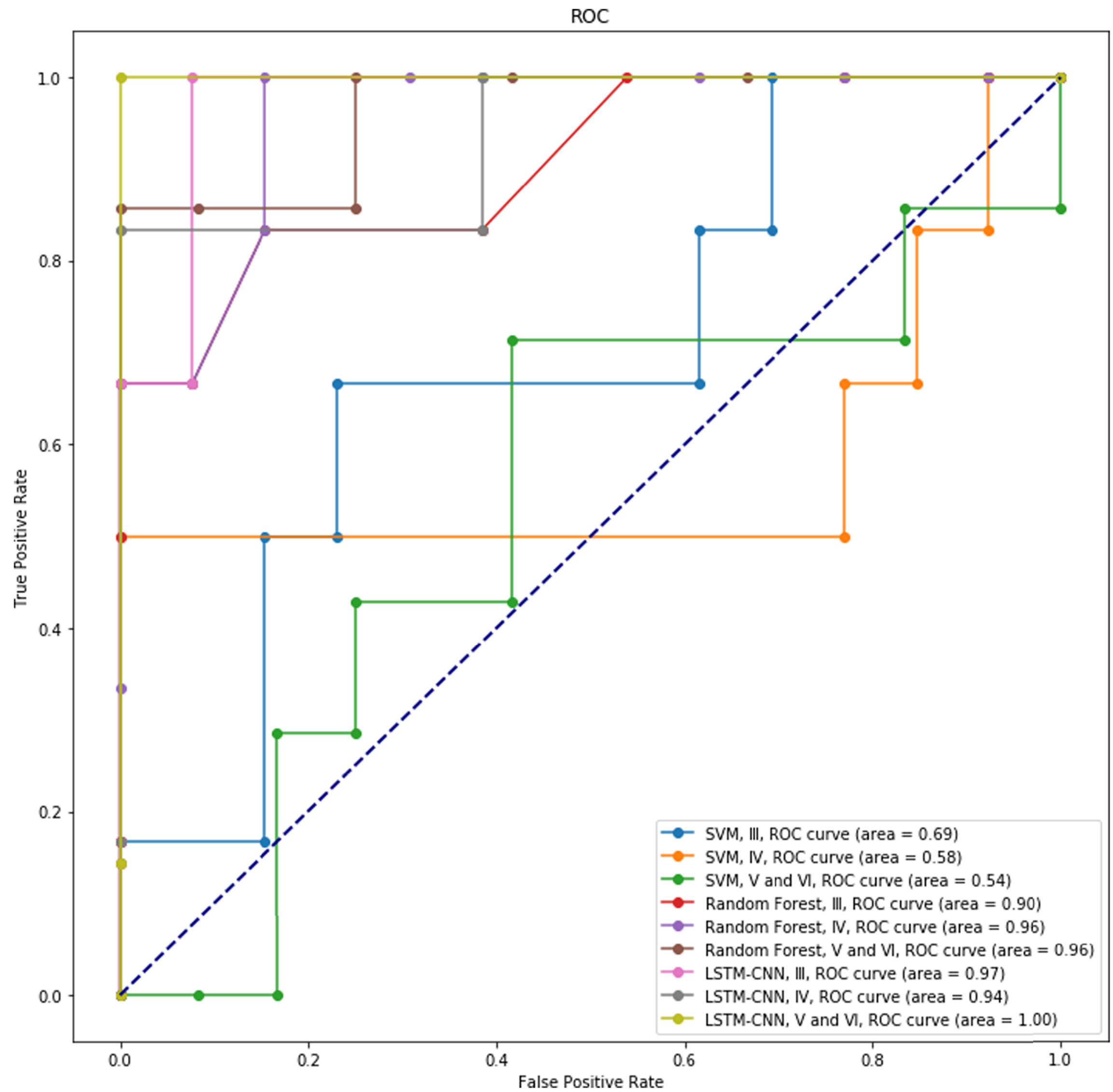

FIGURE 3. The ROC curves of each class for three classification models.

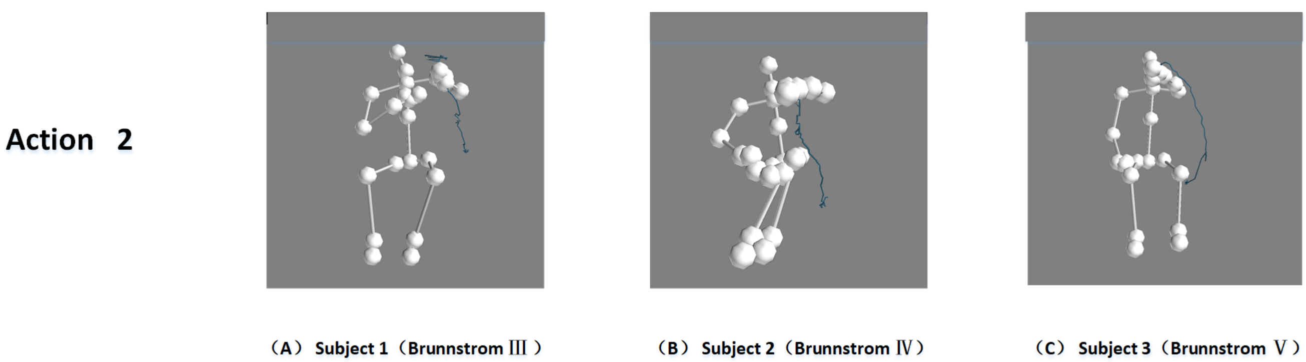

FIGURE 4. The action trajectories of Action 2 from 3-class Brunnstrom stages. The time consumptions were 11.32 s, 10.53 s and 5.16 s, respectively.

for the impaired subjects at home. Finally, we hope that this technology will be adapted to home-based rehabilitation for helping patients who lack professional treatment.

Moreover, our proposed hybrid model was a supervised learning method. That meant that the robustness of the model was supposed to be checked for a mass of clinical trials.

In the future, we will continue to perform the evaluation tests for enough cases. And the stability of our assessment model will be upgraded to the practical level.

V. CONCLUSION

In this paper, we presented a novel hybrid deep network combined LSTM and CNN for rehabilitation assessment referred to Brunnstrom Scale. The experimental result validated the reliability of our proposed evaluation method. And the comparison result of three machine learning algorithms indicated that the superiority of our hybrid model for Kinect-based 3D data analysis.

REFERENCES

[1] P. Langhorne, J. Bernhardt, and G. Kwakkel, ''Stroke rehabilitation,'' *Lancet*, vol. 377, no. 9778, pp. 1693–1702, 2011.

[2] J. Chen and M. I. Shifman, ''Inhibition of neogenin promotes neuronal survival and improved behavior recovery after spinal cord injury,'' *Neuroscience*, vol. 408, pp. 430–447, Jun. 2019.

[3] W. Gabriele and S. Renate, ''Work loss following stroke,'' *Disab. Rehabil.*, vol. 31, no. 18, pp. 1487–1493, 2009.

[4] L. A. Harvey, ''Physiotherapy rehabilitation for people with spinal cord injuries,'' *J. Physiotherapy*, vol. 62, no. 1, pp. 4–11, 2016.

[5] T. Sakaki, N. Ushimi, K. Aoki, K. Fujiie, R. Katamoto, A. Sugyo, and Y. Kihara, ''Rehabilitation robots assisting in walking training for SCI patient,'' in *Proc. IEEE RO-MAN*, Aug. 2013, pp. 814–819.

[6] A. R. Fugl-Meyer, L. Jääskö, I. Leyman, S. Olsson, and S. Steglind, ''The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance,'' *Scand. J. Rehabil. Med.*, vol. 7, no. 1, pp. 13–31, 1975.

[7] J. Chu, L. A. Harvey, M. Ben, J. Batty, A. Avis, and R. Adams, ''Physical therapists' ability to predict future mobility after spinal cord injury,'' *J. Neurolog. Phys. Therapy*, vol. 36, no. 1, pp. 3–7, 2012.

[8] L. A. Harvey, C.-W. C. Lin, J. V. Glinsky, and A. De Wolf, ''The effectiveness of physical interventions for people with spinal cord injuries: A systematic review,'' *Spinal Cord*, vol. 47, no. 3, pp. 184–195, 2009.

[9] X. Huang, F. Naghdy, G. Naghdy, H. Du, and C. Todd, ''The combined effects of adaptive control and virtual reality on robot-assisted fine hand motion rehabilitation in chronic stroke patients: A case study,'' *J. Stroke Cerebrovascular Diseases*, vol. 27, no. 1, pp. 221–228, 2018.

[10] M. Zyda, ''From visual simulation to virtual reality to games,'' *Computer*, vol. 38, no. 9, pp. 25–32, Sep. 2005.

[11] G. Burdea, G. Burdea, N. Langrana, D. Gomez, and B. Liu, ''A virtual reality-based system for hand diagnosis and rehabilitation,'' *Teleoperators Virtual Environ.*, vol. 6, no. 2, pp. 229–240, 1997.

[12] K. H. Cho, K. J. Lee, and C. H. Song, ''Virtual-reality balance training with a video-game system improves dynamic balance in chronic stroke patients,'' *Tohoku J. Exp. Med.*, vol. 228, no. 1, pp. 69–74, 2012.

[13] J. H. Kim, S. H. Jang, C. S. Kim, J. H. Jung, and J. H. You, ''Use of virtual reality to enhance balance and ambulation in chronic stroke: A double-blind, randomized controlled study,'' *Amer. J. Phys. Med. Rehabil.*, vol. 88, no. 9, pp. 693–701, 2009.

[14] A. A. Rendon, E. B. Lohman, D. Thorpe, E. G. Johnson, E. Medina, and B. Bradley, ''The effect of virtual reality gaming on dynamic balance in older adults,'' *Age Ageing*, vol. 41, no. 4, pp. 549–552, 2012.

[15] K. H. Cho and W. H. Lee, ''Virtual walking training program using a real-world video recording for patients with chronic stroke: A pilot study,'' *Amer. J. Phys. Med. Rehabil.*, vol. 92, no. 5, pp. 371–384, 2013.

[16] Y.-R. Yang, M.-P. Tsai, T.-Y. Chuang, W.-H. Sung, and R.-Y. Wang, ''Virtual reality-based training improves community ambulation in individuals with stroke: A randomized controlled trial,'' *Gait Posture*, vol. 28, no. 2, pp. 201–206, 2008.

[17] J. E. Deutsch and A. Mirelman, ''Virtual reality-based approaches to enable walking for people poststroke,'' *Topics Stroke Rehabil.*, vol. 14, no. 6, pp. 45–53, 2007.

[18] M. Brien and H. Sveistrup, ''An intensive virtual reality program improves functional balance and mobility of adolescents with cerebral palsy,'' *Pediatric Phys. Therapy*, vol. 23, no. 3, pp. 258–266, 2011.

[19] D. McEwen, A. Taillon-Hobson, M. Bilodeau, H. Sveistrup, and H. Finestone, ''Virtual reality exercise improves mobility after stroke: An inpatient randomized controlled trial,'' *Stroke*, vol. 45, no. 6, pp. 1853–1855, 2014.

[20] M. Pedraza-Hues, S. Martín-Calzón, F. J. Díaz-Pernas, and M. Martínez-Zarzuela, ''Rehabilitation using Kinect-based games and virtual reality,'' *Procedia Comput. Sci.*, vol. 75, pp. 161–168, Jan. 2015.

[21] K. Khoshelham and S. O. Elberink, ''Accuracy and resolution of Kinect depth data for indoor mapping applications,'' *Sensors*, vol. 12, no. 2, pp. 1437–1454, Feb. 2012.

[22] Z. Zhang, ''Microsoft Kinect sensor and its effect,'' *IEEE MultiMedia*, vol. 19, no. 2, pp. 4–10, Feb. 2012.

[23] J. MacCormick, ''How does the Kinect work?'' *Presentert Ved Dickinson College*, vol. 6, pp. 1–44, Dec. 2011.

[24] J. Preis, M. Kessel, M. Werner, and C. Linnhoff-Popien, ''Gait recognition with Kinect,'' in *Proc. 1st Int. Workshop Kinect Pervasive Comput.*, Jun. 2012, pp. 1–4.

[25] Y. J. Chang, W. Y. Han, and Y. C. Tsai, ''A Kinect-based upper limb rehabilitation system to assist people with cerebral palsy,'' *Res. Develop. Disabilities*, vol. 34, no. 11, pp. 3654–3659, Nov. 2013.

[26] A. Becker, ''Artificial intelligence in medicine: What is it doing for us today?'' *Health Policy Technol.*, vol. 8, no. 2, pp. 198–205, 2019.

[27] S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, ''Can machine-learning improve cardiovascular risk prediction using routine clinical data?'' *PloS ONE*, vol. 12, no. 4, 2017, Art. no. e0174944.

[28] A. Rajkomar, ''Scalable and accurate deep learning with electronic health records,'' *NPJ Digit. Med.*, vol. 1, no. 1, 2018, Art. no. 18.

[29] H. Sak, A. Senior, and F. Beaufays, ''Long short-term memory recurrent neural network architectures for large scale acoustic modeling,'' in *Proc. 15th Annu. Conf. Int. Speech Commun. Assoc.*, Jan. 2014, pp. 338–342.

[30] A. Graves, N. Jaitly, and A.-R. Mohamed, ''Hybrid speech recognition with deep bidirectional LSTM,'' in *Proc. IEEE Workshop Autom. Speech Recognit. Understand.*, Dec. 2013, pp. 273–278.

[31] X. Ma and E. Hovy, ''End-to-end sequence labeling via bi-directional LSTM-CNNS-CRF,'' Mar. 2016, *arXiv:1603.01354*. [Online]. Available: https://arxiv.org/abs/1603.01354

[32] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, ''Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,'' *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[33] Y. LeCun and Y. Bengio, ''Convolutional networks for images, speech, and time series,'' in *The Handbook of Brain Theory and Neural Networks*, vol. 3361, no. 10. Cambridge, MA, USA: MIT Press, 1995.

[34] O. Abdel-Hamid, A.-R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, ''Convolutional neural networks for speech recognition,'' *IEEE/ACM Trans. Audio, Speech Language Process.*, vol. 22, no. 10, pp. 1533–1545, Oct. 2015.

[35] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ''ImageNet classification with deep convolutional neural networks,'' in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[36] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, ''MobileNets: Efficient convolutional neural networks for mobile vision applications,'' Apr. 2017, *arXiv:1704.04861*. [Online]. Available: https://arxiv.org/abs/1704.04861

[37] T. Young, D. Hazarika, S. Poria, and E. Cambria, ''Recent trends in deep learning based natural language processing,'' *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, Aug. 2018.

[38] C.-W. Hsu and C.-J. Lin, ''A comparison of methods for multiclass support vector machines,'' *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

[39] A. Subudhi, M. Dash, and S. Sabut, ''Automated segmentation and classification of brain stroke using expectation-maximization and random forest classifier,'' *Biocybern. Biomed. Eng.*, to be published.

[40] J. Sarsfield, D. Brown, N. Sherkat, C. Langensiepen, J. Lewis, M. Taheri, C. McCollin, C. Barnett, L. Selwood, P. Standen, P. Logan, C. Simcox, C. Killick, and E. Hughes, ''Clinical assessment of depth sensor based pose estimation algorithms for technology supervised rehabilitation applications,'' *Int. J. Med. Inform.*, vol. 121, pp. 30–38, Jan. 2019.

[41] M. Capecci, M. G. Ceravolo, F. Ferracuti, S. Iarlori, V. Kyrki, S. Longhi, L. Romeo, and F. Verdini, ''Physical rehabilitation exercises assessment based on hidden Semi-Markov model by Kinect v2,'' in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Inform. (BHI)*, Feb. 2016, pp. 256–259.

[42] N. Kitsunezaki, E. Adachi, T. Masuda, and J.-I. Mizusawa, "KINECT applications for the physical rehabilitation," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, vol. 43, May 2013, pp. 294–299.
[43] D. Leightley, J. Darby, B. Li, J. S. McPhee, and M. H. Yap, "Human activity recognition for physical rehabilitation," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2013, pp. 261–266.

LEI CAO received the Ph.D. degree in computer software and theory from Tongji University, China, in 2016. He was a Lecturer with the Shanghai Maritime University, China. He was a Visit Researcher with the Institute of Medical Psychology and Behavioral Neurobiology, Tuebingen University, in 2013. He has authored or coauthored about 20 journal and conference papers. His research interests include EEG signal processing, brain–computer interface, machine learning, and medical image analysis.

CHUNJIANG FAN received the master's degree in medical rehabilitation from Jiangnan University, China. He is currently pursuing the Ph.D. degree with Suzhou University. His research interests include medical image analysis and cognitive neuroscience.

HAORAN WANG received the B.S. degree in computer science from Shanghai Maritime University, China. He is currently pursuing the Ph.D. degree with Tongji University. His research interests include machine learning, medical image analysis, and brain–computer interface.

GUOKAI ZHANG is currently pursuing the Ph.D. degree with the College of Software Engineering, Tongji University, Shanghai, China. His current research interests include deep learning, object detection, and medical image analysis.