Deep-Learning-Based Emergency Stop Prediction for Robotic Lower-Limb Rehabilitation Training Systems

Baekdong Cha, Kyung-Hwan Lee, and Jeha Ryu, *Member, IEEE*

***Abstract*— Robotic lower-limb rehabilitation training is a better alternative for the physical training efforts of a therapist due to advantages, such as intensive repetitive motions, economical therapy, and quantitative assessment of the level of motor recovery through the measurement of force and movement patterns. However, in actual robotic rehabilitation training, emergency stops occur frequently to prevent injury to patients. However, frequent stopping is a waste of time and resources of both therapists and patients. Therefore, early detection of emergency stops in real-time is essential to take appropriate actions. In this paper, we propose a novel deep-learning-based technique for detecting emergency stops as early as possible. First, a bidirectional long short-term memory prediction model was trained using only the normal joint data collected from a real robotic training system. Next, a real-time threshold-based algorithm was developed with cumulative error. The experimental results revealed a precision of 0.94, recall of 0.93, and F1 score of 0.93. Additionally, it was observed that the prediction model was robust for variations in measurement noise.**

***Index Terms*— Deep learning, emergency stop prediction, robotic lower-limb rehabilitation, time series prediction.**

I. Introduction

THE goal of rehabilitation exercises is to perform specific movements that induce motor plasticity to improve motor recovery and minimize functional deficits. Modern concepts of motor learning favor task-specific trainings such as relearning walking. However, they require significant physical effort from the physiotherapists to assist the patients and additional training for free walking with guidance by at least two physiotherapists [1].

Robotic rehabilitation may be a solution for automated training [2]. This technology can provide accurate proprioceptive, kinematic, and kinetic guidance, as well as variable error practice, high-intensity, and repetitive task-specific and interactive exercises for paretic lower-limbs [3], [4]. Based on these benefits, exoskeleton-type robotic lower-limb rehabilitation systems are used in medical fields [3], [5]–[8] resulting in substantially improved lower-limb function.

For robotic training, pre-tests are performed on all patients to identify the ones with conditions that complicate the rehabilitation robot training [9]. Setting the correct training parameters (e.g., walking speed and weight support ratio) can enhance the training efficacy and prevent potential safety problems that could occur during the training. Despite these precautionary measures, emergency stops tend to occur frequently in real rehabilitation training with exoskeleton-type robotic systems (e.g., Walkbot (P&S Mechanics, Korea) and Lokomat (Hocoma, Switzerland)) to avoid injury to the patients. One of the major causes of emergency stops is spasticity, which results in an increased level of joint resistance during passive movement, because repeated passive lower-limb exercises are characteristic of lower-limb rehabilitation robots. Other causes may include incorrect settings by inexperienced physical therapists, improper wearing of the rehabilitation robot, and/or awkwardness of body movements in patients during the robot rehabilitation training for the first time. Such emergency stops are also likely to occur in other wearable robotic walk-assist devices for healthy individuals. To date there has been no comprehensive study of abnormal events such as emergency stops during the robot rehabilitation training. Our clinical training experiences and those shared by others indicate that emergency stops occur frequently in rehabilitation robot systems.

Emergency stops are a waste of time and resources of both therapists and patients. A therapist typically monitors a patient's measured kinematic leg and knee angles and estimated kinetic leg and knee torques [10] reported by the robot's sensors during the training to determine whether the patient is correctly undergoing the training. However, the complexity of wearable robots including the lower-limb rehabilitation robots, makes it difficult to identify abnormal events in real-time based solely on the trends of time series data. Therefore, it is necessary to develop an intelligent and automated method to

Manuscript received February 9, 2021; revised April 12, 2021 and May 4, 2021; accepted June 6, 2021. Date of publication June 9, 2021; date of current version June 17, 2021. This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korean Government (MSIT) (Development of artificial intelligence technology for smart lower-limb rehabilitation) under Grant 2020-0-01129. *(Corresponding author: Jeha Ryu.)*

Baekdong Cha is with the School of Integrated Technology, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, Republic of Korea (e-mail: orientpine@gmail.com).

Kyung-Hwan Lee is with P&S Mechanics Company Ltd., Seoul 07294, Republic of Korea (e-mail: leeunderkino@gmail.com).

Jeha Ryu is with the School of Integrated Technology, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, Republic of Korea, and also with the Artificial Intelligence Graduate School Program (GIST), Gwangju Institute of Science and Technology (GIST), Gwangju 61005, Republic of Korea (e-mail: ryu@gist.ac.kr).

Digital Object Identifier 10.1109/TNSRE.2021.3087725

predict improper operations such as emergency stops, improve the efficacy of the rehabilitation process, and avoid excessive emergency stops during the training.

The scenario for resolving emergency stops that may occur during the operation of a lower-limb rehabilitation robot consists of three steps: 1) predicting the emergency stops, 2) identifying the causes of emergency stops, and 3) taking appropriate actions to eliminate the causes of emergency stops in real-time before the stops occur. In this study, we have focused only on predicting the emergency stops as early as possible in real-time. Steps 2 and 3 require significant analysis and are planned for future research.

In related studies, Arami *et al.* [11] analyzed the machine learning-based prediction of gait freezing in Parkinsonian patients based on recent studies using a set of possible markers of the freezing of gate (FOG) [12], including cadence increase, decrease in step length, appearance of new peaks, and higher frequency of acceleration during and/or prior to FOG. A similar approach in [13] hypothesized the degeneration of gait patterns prior to FOG. This approach could be applied to the prediction of emergency stops in robotic rehabilitation systems and in passive training where a patient's lower-limbs are fixed to the robot's legs and the patients are instructed to be passive while being guided by the robot to walk. However, this machine learning-based method requires the following subtasks: 1) a set of movement-based features should be defined and identified to accurately distinguish the anomalies from the normal rehabilitation training using handcrafted machine learning (ML) approaches and 2) a subset of these features should be effectively predicted based on past values using a data-driven predictive model. The predicted features can then be used to predict the specific anomalies. These subtasks require large amount of data that must be collected using an appropriate design with a sufficient number of diverse patients; this requires significant time and effort. This type of handcrafted ML approach dealing with time series data generally requires good domain knowledge; hence, it is difficult to establish optimal hyperparameters for optimizing the model. Moreover, in the rehabilitation training system, there are many parameters (e.g., walking speed, body weight support, and force guidance) related to the therapy protocol (e.g., when it is necessary to adjust the walking speed according to the patient's condition) thus, complicating the application of handcrafted ML approaches. However, the deep learning (DL) approaches are more appropriate because the DL model can extract the essential features implicitly without any prior domain knowledge.

Real datasets have already been collected during real patient clinical trainings using the Walkbot (P&S Mechanics, Korea) over the past ten years. The dataset contains anonymized log data for the system maintenance of Walkbot. These datasets include the history of 69 patient trainings with sudden stop events due to unknown causes. Although these datasets were not collected using an appropriate design(we do not know the reasons for the occurrence of emergency stops), they can be used to predict emergency stops in advance because we were able to identify abnormal behaviors prior to the emergency stops in most cases (approximately 90% of the collected dataset). The three typical data presented in Fig. 1 represent the abnormal behaviors before an emergency stop.

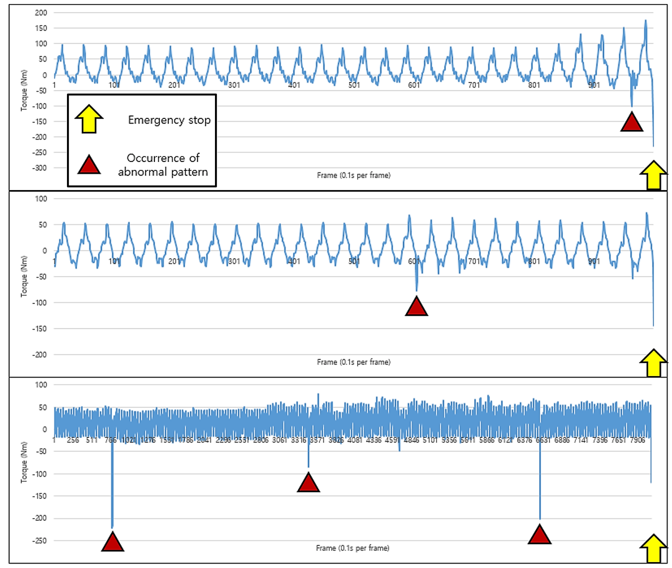

Fig. 1. Typical data showing abnormal behaviors before an emergency stop: (a) top, (b) middle, and (c) bottom.

These figures approximately indicate the regular hip torque patterns for a certain duration after the walking training starts and then begin to exhibit certain abnormal patterns, such as an increasing negative magnitude beyond a typical regular value and one (Figs. 1(a) and 1(b)) or more (Fig. 1(c)) very large negative amplitude (indicated by red triangles) immediately before (one stride before in Fig. 1(a)) or long before (11 strides before in Fig. 1(b) and multiple strides before in Fig. 1(c)), which is an emergency stop (indicated by yellow arrows). This typical behavior suggests that a method should be used for predicting an emergency stop before it occurs. The potential start (indicated by red triangles) of abnormal behavior was determined by a group of physical therapists who specialized in rehabilitation of lower-limb using rehabilitation robots.

For datasets in which it is difficult to analyze the characteristics of emergency stops and the features of datasets, DL-based anomaly detection methods can be applied [14]. In such methods, a prediction model can be trained using a sufficient number of normal data that are readily available from non-emergency stop cases or prior to abnormal behaviors. Once a prediction model is trained, the future value of the test data can be predicted and compared to the ground-truth data that are collected in real-time during the training. If the error between the predicted and ground-truth values is larger than a predefined threshold, then it is highly likely that an emergency stop will occur in the near future.

This paper proposes a DL-based real-time emergency stop prediction method that can be used in robotic gait rehabilitation training systems without prior anomaly knowledge. The DL-based model learns and predicts the pattern of normal data; based on this result, a threshold-based algorithm is developed to predict an emergency stop when an abnormal behavior is observed. The proposed method uses real datasets collected during real patient clinical trainings using the Walkbot (P&S Mechanics, Seoul, Korea). The DL module is trained

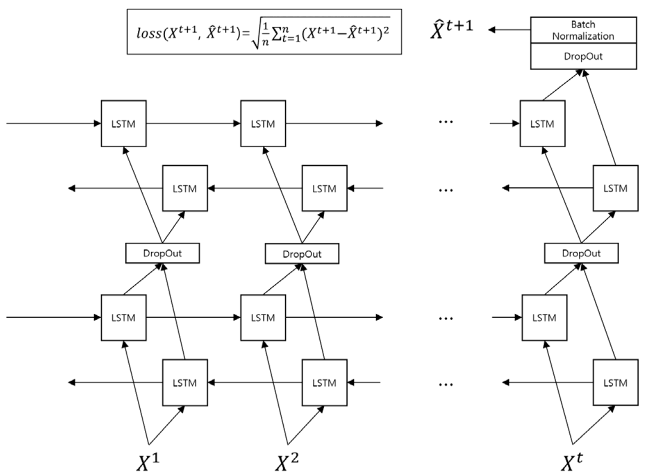

Fig. 2. Modified two-layered bidirectional long short-term memory (bi-LSTM) recurrent network.

using a bidirectional long short-term memory (bi-LSTM) [15] that requires the history of normal data without sudden stops. When predicting emergency stops, a full dataset containing the histories of sudden stops was used. A dataset that did not include emergency stops was also used to verify the performance of the model. The proposed method predicted the emergency stops before they occurred (approximately 0.5 s). This early detection is sufficient for taking preventive measures, such as increasing weight support, reducing walking speed, and reducing stride before an emergency stop occurs, considering the 0.1 s control rate of the lower-limb rehabilitation training system.

The main contributions of this study are twofold. First, a fast and accurate emergency stop prediction method is proposed for robotic lower-limb rehabilitation training systems using bi-LSTM, which is the first of its kind. Second, the method was evaluated using real patient datasets obtained from the rehabilitation training.

II. Prediction of Emergency Stops

The datasets used in this study consist of time-series data. Hence, a recursive neural network was adopted to solve the prediction problem. This network consists of two components: (i) predicting the value of the next time instance in the time series data and (ii) detecting anomalies such as incoming emergency stops by comparing predicted values to real values.

A. Future Prediction

In this study, we adopted the two-layered bi-LSTM recurrent network [16] as shown in Fig. 2, which was developed for repetitive cyclic events. This network uses both past (forward) and future (backward) gait cycle states and dropout techniques to reduce the network size and avoid overfitting. The goal of stacking multiple LSTM models in such a multi-layered hierarchical architecture is to extract features in the lower layers that will disentangle the factors of variation in the input data and then combine these features in higher layers.

The authors of [15] used the bi-LSTM architecture to classify normal and pathological gaits (knee rigidity and limping) because, the lower-limb joint angles in the current frame are closely related to both the previous and future frame angles in a walking sequence; additionally, one joint angle is related to other joint angles. This architecture can be applied to arbitrary joint features, namely kinematic (e.g., angle) or kinetic (e.g., torque) because kinematic and kinetic data are correlated [17]. It can also be applied in our study to predict the future joint features (hip torque) using previous feature data.

The two-layered bi-LSTM recurrent neural network shown in Fig. 2 is a modified version of that presented in [15]. The inputs are the previous joint features X^t up to the current time t and the output is the predicted state at the next time $(t + 1)$. The predicted output is compared with the ground-truth data to calculate a loss function, and backpropagation process is used to minimize the loss by changing the network weights. For the regression task, we selected the root mean square error (RMSE) [18] loss function to improve the training and validation accuracy of the model. In the modified network outputs, batch normalization was applied to increase the training speed of the model [19]. Each batch of inputs for the model (and other feed-forward activations) have different statistics that are not typically representative of the training data as a whole. Therefore, the layers of the model must constantly adapt to the changing statistics resulting in inefficient training. However, from the history of input data only the hip torque data in the sagittal plane were used in the prediction problem because the collected dataset revealed that most emergency stops were associated with kinetic hip torque data. The number of emergency stops caused by the knee were relatively small, approximately 5%, compared to the emergency stops caused by the hip. Therefore, other kinematic hip and knee angles and kinetic knee torques were not used.

B. Emergency Stop Detection

Once a future sample (x_{t+1}, hip torque) is predicted by the bi-LSTM, it is compared with the corresponding ground-truth data sample (y_{t+1}) to define the prediction error E_{t+1} as follows:

$$E_{t+1} = |y_{t+1} - x_{t+1}| \,. \tag{1}$$

The prediction errors in the hip torque data histories can be used directly to detect the upcoming emergency stops. For example, an emergency stop may occur if the single-step prediction error defined in (1) (the so-called instantaneous error) in one future frame moves outside a predefined confidence interval. This confidence interval can be set as the range surrounding the ground-truth data sample X (X_{max} and X_{min}) of a patient's hip torque during a particular time interval (approximately three cycles of hip torque data) and is defined as:

$$Confidence\ interval = (X_{max} - X_{min}) \times r, \tag{2}$$

where r is a pre-selected ratio between zero and one. Fig. 3 illustrates the true and predicted values and confidence intervals (shaded area) with two cases of prediction errors. Prediction error 1 is small; hence, it is estimated to be normal, whereas Prediction 2 is sufficiently large to be estimated as

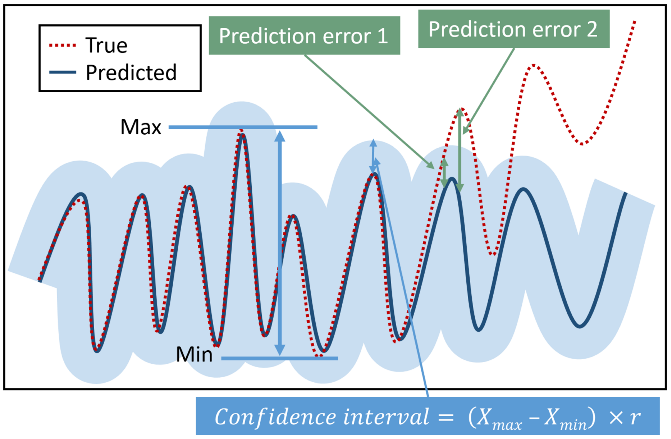

Fig. 3. Illustration of confidence intervals and prediction errors.

abnormal (i.e., an emergency stop may occur in the near future).

A large confidence interval may not detect true emergency stops, while a small confidence interval may detect too many emergency stops (false alarms). Additionally, noise in the signal can generate incorrect judgments particularly for small confidence intervals. Smaller the value of r, greater is the effect of noise. To avoid these issues, another type of error called cumulative error (CE) c_t is used. CE accumulates the previous prediction errors across multiple frames over a short period of time (one or more frames depending on the data) prior to signaling a sustained anomaly. CE can reduce false positive rates compared to an instantaneous error for one future frame.

When computing the CE, two factors should be considered: first, CE should be computed only if the prediction error is outside the confidence interval defined in (2); otherwise, it should be set to zero. If the CE accumulates prediction errors, which are positive values, without resetting to zero, it can increase without indicating an anomaly. If only prediction errors outside the confidence interval are accumulated, then the CE can indicate anomalies more clearly.

Fig. 4 presents a typical abnormal behavior in which high negative peaks are clearly visible just before an emergency stop (see the magnified view at the bottom). This figure shows that while the instantaneous error (black solid line) increases gradually, the CE (yellow solid line) exhibits sharp changes but resets to zero if the instantaneous error is within the confidence interval. These sharp changes are easier to detect than gradual changes. This figure also presents the actual emergency stop frame (thick red arrow) and frame (thick yellow arrow) at which the emergency stop is predicted in advance by the proposed method. The thick green arrow represents the detection horizon indicating the early detection of an emergency stop.

The second factor related to CE is to consider the change in predicted and true values between two consecutive time frames. Fig. 5 presents seven typical cases of predicted (blue line) and ground-truth (true) (red line) values from $(t-1)$ to (t). Case 1: same signs and converging, Case 2: same signs and diverging, Case 3: same signs and crossing, Case 4: same signs and parallel, Case 5: different signs and diverging, Case 6: different signs and converging, and Case 7: different signs and crossing.

Algorithm 1 Computation of CE

1: **PROCEDURE**: Determine when the prediction error E_t is accumulated.

2: Input: x_t, y_t, q, T, r; True joint data: x_t; Predicted joint data: y_t; Number of past data (fixed length): q; Particular time instance: T ($>q+1$); ratio (threshold for confidence interval): r

3: Output: cumulative error: c_t

4: Initialize $c_T \leftarrow 0$

5: **repeat**

6: $t \leftarrow T, X = (x_{T-q}, x_{T-q+1}, x_{T-q+2}, \ldots, x_{T-1})$;

5: Calculate X_{max}, X_{min};

6: **if** $E_T > (X_{max} - X_{min}) \times r$ **and** $(x_{T-1} - y_{T-1}) \times (x_T - y_T) > 0$ **then**

7: $c_T \leftarrow c_T + E_T$;

8: **Else**

9: $c_T \leftarrow 0$;

11: **return** c_T;

12: $T \leftarrow T + 1$;

13: **until** emergency stop is detected;

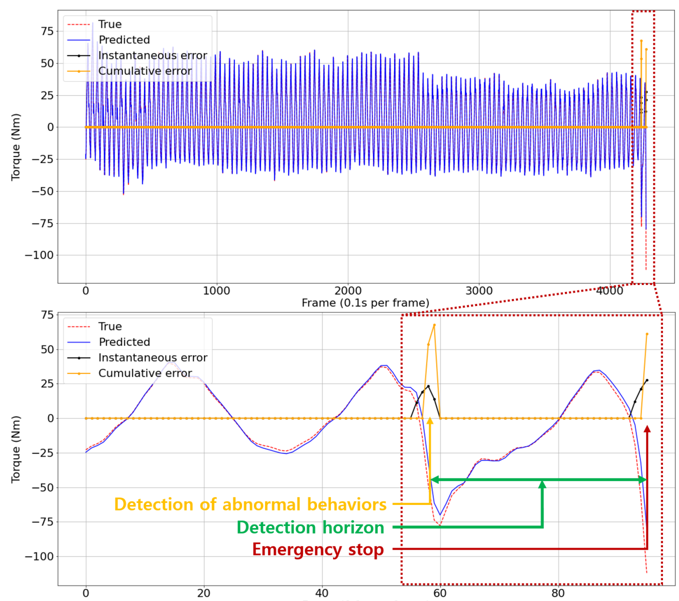

Fig. 4. Typical error behaviors before emergency stops: the bottom figure presents a magnified view before an emergency stop.

Fig. 5 demonstrates that the predicted values may cross over the true values (see Cases 3 and 7) between two consecutive time frames, which can be checked using the following inequality condition:

$$Cross\ over\ happens\ if\ \ (x_{T-1}-y_{T-1})\,(x_{T-1}-y_{T-1}) < 0. \tag{3}$$

When computing the CE, accumulation of all the predicted errors may not reveal sharp CE changes. Hence, to capture sharp changes, we suggest that crossover cases should be

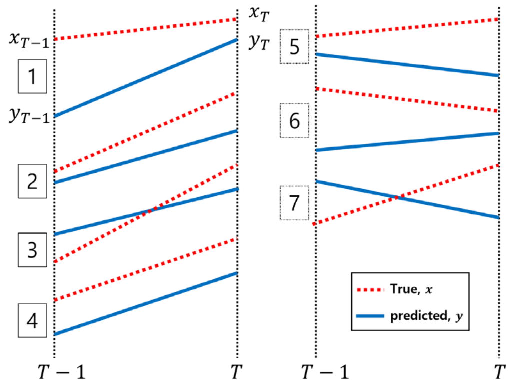

Fig. 5. Predicted and ground-truth (true) values between two consecutive time frames.

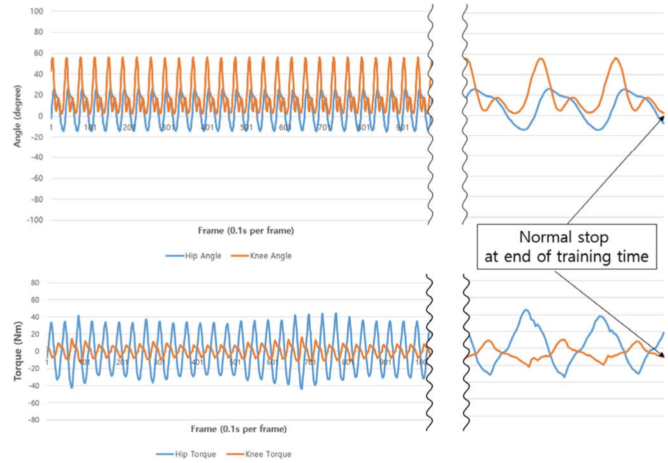

Fig. 6. Typical normal joint data.

excluded when computing the CE. Accordingly, we tested two cases (with and without the inclusion of crossover cases) and determined that excluding the crossover cases resulted in a better performance as discussed in Section IV.

The following summarizes the proposed CE computation method.

Fig. 4 presents a typical CE behavior (sharp change just before an emergency stop) based on Algorithm 1 for the data of one patient. This result demonstrates that the proposed CE method can easily and accurately detect emergency stops in advance.

Once the emergency stops are predicted by training the bi-LSTM network on large amount of normal data (i.e., hip torque data without emergency stops), the proposed method is implemented using the following steps:

(i) Collect joint data (hip torque in this study) from a patient using the robot rehabilitation system in real-time.

(ii) Input the previously collected data q into the neural network. In this study, we used $q = 108$, which is sufficiently large (approximately three cycles of hip torque data), to compute the maximum and minimum values of the time-sequence data for confidence interval estimation. This implies that the proposed method can only be applied after the collection of hip torque data.

(iii) Predict the future joint data at the next time step $(t+1)$. The proposed algorithm does not require any prior knowledge of abnormal behaviors and utilizes normal datasets for training.

(iv) Compute the prediction error using (1).

(v) Identify anomalous behaviors using the CE computed by Algorithm 1. If the CE is greater than a predefined threshold, then an abnormal behavior is identified.

III. Experiments

A. Data

The gait training data used in this study were collected from 69 patients trained by Walkbot (P&S Mechanics, Seoul, Korea). Most of the patients were diagnosed with walking disorders, such as stroke, cerebral palsy, and Parkinson's disease. Because the dataset contains anonymized log data, there is no particular dataset population. However, the population consisted of mostly stroke patients (approximately 80%).

Several sessions of gait training data were gathered from each patient. The collected data were classified by the gait training physiotherapists as normal or abnormal patterns (149 normal sessions from 31 patients and 67 abnormal sessions from 38 patients). Each session dataset constituted the angle and torque trajectories of the two joints (hip and knee). Each session of normal termination lasted for approximately 30 min. In the event of an emergency stop, the session length depended on the time of occurrence of abnormal behaviors.

Note that abnormal behaviors were not observed before emergency stops while collecting the dataset. Additionally, physical therapists did not notice any abnormality in the patients or the machine Walkbot. Emergency stops occur when the safety protection of the hip joint exceeds the threshold. Hence, the Walkbot is programmed to induce an emergency stop if the hip joint torque exceeds the preset threshold for preventing injury to patients.

Fig. 6 presents the representative histories for normal training sessions; they are relatively consistent and exhibit no significant changes in joint data values throughout the training duration. A typical emergency stop during the training is presented in Fig. 7, where an abnormal pattern appears in the torque data prior to the emergency stop. However, the joint angle data did not reveal any abnormal behaviors because the training data were collected during passive modes of lower-limb rehabilitation training where the patient's lower-limbs were fixed to the robot's legs. This means that the trajectories of the joints only follow the specified gait patterns designed by physiotherapists for proper rehabilitation training for each patient. Therefore, abnormal behavior detection should be performed based on the history of torque data of the joints.

B. Data Preprocessing

The raw data from robotic rehabilitation systems are noisy; hence, a fourth-order Butterworth filter with a cut-off frequency of 5 Hz was used to filter out random noise and retain meaningful signals representing abnormal behaviors [20]. Additionally, data normalization was performed to ensure that the values of input data were between zero and one.

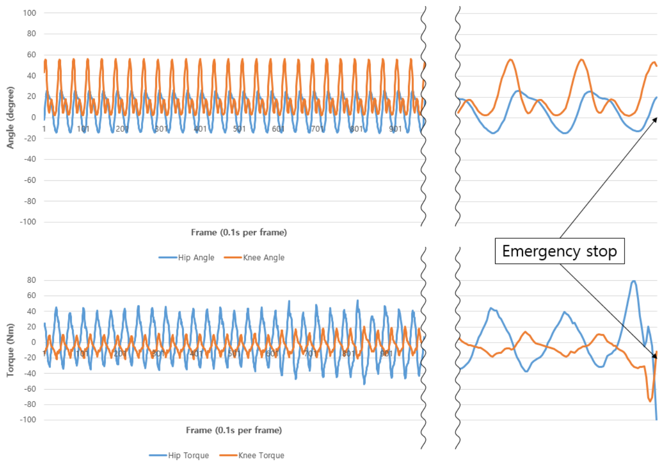

Fig. 7. Typical abnormal joint data.

To train the bi-LSTM network for future joint data prediction, 82 normal datasets for 17 patients were used and data selection was randomly sampled. The 82 training datasets were split into 8:2 (training : validation). The evaluation of model performance in detecting abnormal behaviors used a total of 134 testing datasets (67 normal datasets from 14 patients that were not used for training and 67 abnormal datasets from 38 patients). Therefore, the individuals in the training set were not present in the testing set. For reliable metrics, such as precision and recall a similar number of normal and abnormal datasets must be used.

C. Ablation Study

The proposed model was implemented in KERAS Framework 4 with TensorFlow as the backend. The proposed model and algorithm were executed on a machine with an AMD Ryzen 7 2700X CPU with eight cores and an NVIDIA GeForce GTX 1080ti GPU. The Adam optimizer [21] was used with optimal parameters (learning rate = 0.0012, beta1 = 0.9, beta2 = 0.999, epsilon = 1.0e−08, no decay).

We performed an ablation study to determine the best parameters of the bi-LSTM model by verifying the performance for: (1) the number of layers in the four ranges of [1, 2, 3, 4], (2) number of hidden units in the seven ranges of [5, 10, 30, 50, 100, 200, 300], and (3) number of inputs of the previously collected datasets (q) in the six ranges of [18, 36, 72, 108, 144, 216] in one cycle (36 frames) of walking. This ablation study required a total of 168 ($4 \times 7 \times 6$) cases, and training for each case lasted for approximately 20 min to 2 h depending on the size of the architecture and input data.

For accurate evaluation and optimization of the model hyperparameter, we used 5-fold cross validation. The model stopped training early when the validation loss (RMSE) did not improve for more than 10 epochs to prevent overfitting.

Table I summarizes the RMSE values of the test set. Note that the results in Table I present only the best RMSE for the best number of hidden nodes and input dataset (q) for each number of layers (1, 2, 3, and 4) among a total of 168 cases. Considering the number of weights for inference time (in the order of 0.08 s) and minimum RMSE, we selected the best architecture with two layers, 150 hidden units, and 108 input datasets.

TABLE I
ABLATION STUDY OF BI-LSTM NETWORK

No. of layer	No. of hidden units	No. of q	RMSE
1	30	18	0.0278
2	150	108	0.01989
3	100	72	0.02069
4	100	144	0.0237

D. Performance Evaluations

The proposed algorithm applies two main enhancements: (i) the use of CE instead of instantaneous error and (ii) the consideration of crossing cases in the CE in Algorithm 1 to investigate the effect of these enhancements in the following four cases:

1) AYEY (Accumulator Yes, Exclusion Yes): Use cumulative error, exclude crossing cases.
2) AYEN (Accumulator Yes, Exclusion No): Use cumulative error, include crossing cases.
3) ANEY (Accumulator No, Exclusion Yes): Use instantaneous error, exclude crossing cases.
4) ANEN (Accumulator No, Exclusion No): Use instantaneous error, exclude crossing cases.

The AY (Accumulator Yes) method accumulates E_T if $E_T > (X_{max} - X_{min}) \times r$ and detects abnormal behavior when CE exceeds the emergency stop prediction threshold. The AN (Accumulator No) method detects abnormal behavior when the instantaneous error E_T exceeds the threshold if $E_T > (X_{max} - X_{min}) \times r$ without accumulating E_T. The EY (Exclusion Yes) method computes E_T only if $(x_{T-1} - y_{T-1})(x_{T-1} - y_{T-1}) > 0$. The EN (Exclusion No) method does not compute E_T.

To evaluate the performances of the four cases described above, various cumulative ratios (r) and emergency stop prediction thresholds ($Threshold$) were tested. The cumulative ratios were set to 1, 2, 3, 4, 5, 8, 11, 14, and 20% of the hip torque peak-to-peak amplitude (i.e., r = [0.01, 0.02, 0.03, 0.04, 0.05, 0.08, 0.11, 0.14, 0.17, 0.2]). A range of meaningful cumulative ratios was identified with grid search and the values within that range were set as targets.

The emergency stop prediction thresholds were set to $Threshold$ = [0, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 60, 70, 80, 100, 120, 140, 160, 180, 200]. These thresholds were used for both CE (that requires a high detection threshold because it is the sum of E_T) and instantaneous error (that requires a low detection threshold because it detects abnormal behavior based on a single E_T value).

IV. EXPERIMENTAL RESULTS

To evaluate the proposed emergency stop prediction algorithm, we used metrics of precision, true positive rate (TPR), specificity, false positive rate (FPR), and $F1$ score as discussed in [22].

$$Precision = \frac{T_p}{T_p + F_p}, \tag{4}$$

$$TPR = Recall = Sensitivity = \frac{T_p}{T_p + F_n}, \tag{5}$$

$$Specificity = \frac{T_n}{T_n + Fp}, \tag{6}$$

$$FPR = 1 - Specificity = \frac{F_p}{T_n + F_p}, \tag{7}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{8}$$

where T_p represents the number of correctly detected anomalies (true positive), F_p represents the number of false positives, T_n represents the number of true negatives, and F_n represents the number of false negatives in the confusion matrix. Precision is defined as the ratio of true positives to the total number of positive outputs predicted by the model. Recall is the fraction of positive examples correctly predicted by a model. $F1$ score is the harmonic mean of the precision and recall.

To illustrate the diagnostic capabilities of a binary classifier system as its discrimination threshold varies, a receiver operating characteristic (ROC) curve can be plotted [23]. The area under the ROC curve (AUC) provides an aggregate measure of the performance across all possible classification thresholds. Figs. 8 and 9 presents the ROC curves based on instantaneous error, and ROC curves based on CE, respectively.

The ROC curves reveal the differences in performance based on the parameters of the ratio (r) and emergency stop prediction threshold ($Threshold$) in the proposed algorithm. Hence, we can select optimal values for these two parameters based on the curves and areas. A comparison of the AUCs in Figs. 8 and 9 reveals that CE provides better performance than instantaneous error. The best ratio is $r = 0.11$ in the AYEY condition in Fig. 10, which yields the maximum AUC. The optimal emergency stop prediction threshold ($Threshold$) can be selected from Table II, where the numerical values of various performance measures (threshold, TPR (Recall), FPR, precision, and F1 score) are computed for the best ratio of 0.11 in the AYEY condition. Here, $Threshold = 15$ yields the maximum F1 score.

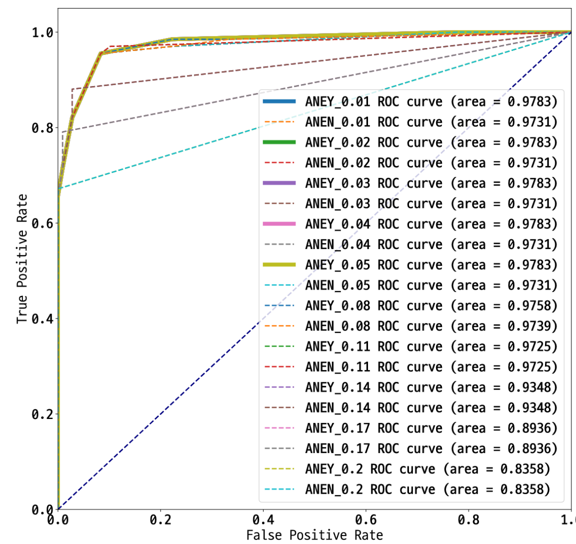

Fig. 8. ROC curves using instantaneous error.

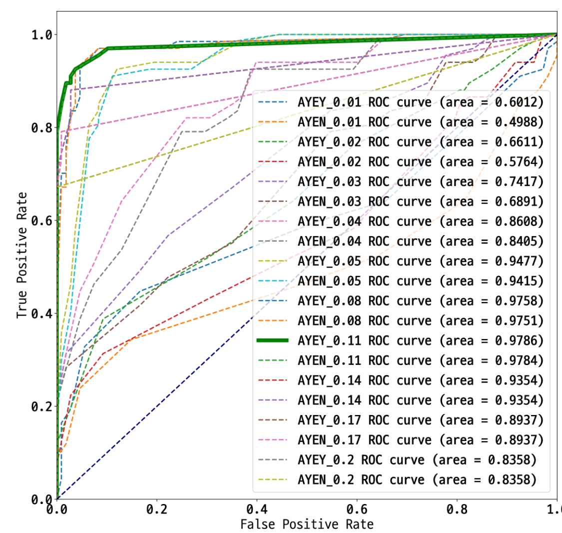

Fig. 9. ROC curves using CE.

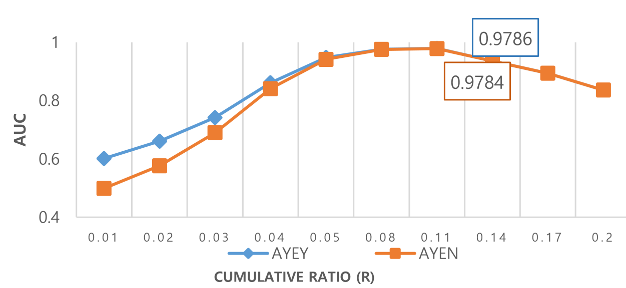

Fig. 10. AUC versus cumulative ratio (r) in the proposed algorithm under AYEY and AYEN conditions.

V. Discussion

Based on the results in the previous section 4, the condition of excluding the crossing case (EY) exhibits a similar or better performance in all cases compared to EN. This condition can reduce the FPR while maintaining the TPR of emergency stop detection because EN in the normal training data causes an increase in the FPR.

For instantaneous error, lower the confidence interval higher is the emergency stop prediction performance. This result mainly stems from multiple trial-and-error efforts to reduce the noise of the joint torque to improve the performance of the prediction model. However, the performance is still vulnerable to noise in signals because, lower the confidence interval more sensitive it is to relatively small value changes. Therefore, instantaneous error is not recommended for the robust prediction of emergency stops. As demonstrated in the experimental results, the AYEY condition yielded good performance with strong robustness against noise in the joint torque data.

TABLE II
TPR (Recall), FPR, Precision and F1-Score Versus Threshold Values in Optimal Algorithm With Optimal Threshold

Threshold	TPR(Recall)	FPR	Precision	F1
0	0.9701	0.1019	0.8553	0.9091
5	0.9701	0.1019	0.8553	0.9091
10	0.9701	0.1019	0.8553	0.9091
15	0.9254	0.0370	0.9394	0.9323
20	0.8955	0.0278	0.9524	0.9231
25	0.8955	0.0185	0.9677	0.9302
30	0.7910	0.0000	1.0000	0.8833
35	0.7612	0.0000	1.0000	0.8644

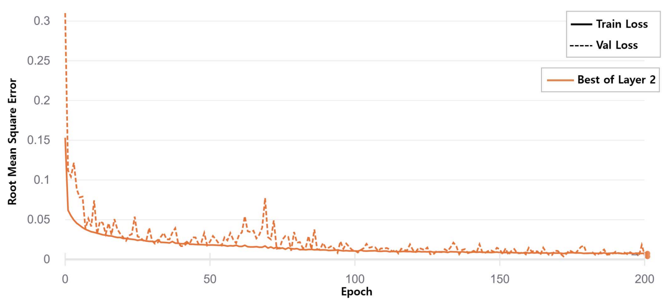

Fig. 11. RMSE in training and validation vs epoch.

Although the proposed emergency stop detection relies on two thresholds (the cumulative ratios (r) and the emergency stop prediction threshold) owing to their sensitivity to the thresholds, they are not excessively sensitive in a certain range (for example, when the cumulative ratios (r) are within 5% to 17% and when the emergency stop prediction threshold is within 10 – 25). Therefore, the proposed method can be generalized to similar problems by choosing optimal threshold values and trade-off between true and false positives.

The dataset was obtained while training patients on a rehabilitation robot system *in passive mode* that only repeats a prescribed walking trajectory (in the sagittal plane for each leg) depending on the patient training protocol. Although gait disorders can be different among diverse populations (and thus high variability), the torque data variability among patients is not large because the patient is trained by repetitive passive motion while tightly coupled with the rehabilitation robot system. Therefore, we believe that the proposed method can be generalized to other rehabilitation robotic systems using passive mode training.

The proposed method predicted the emergency stop using only the hip torque data as a univariate time series. Because the human lower-limb is organically connected, a more accurate prediction may be possible if multivariate time series that include joint data are used.

The emergency stop prediction performance was affected by the network prediction performance. If the value predicted by the network is not sufficiently accurate, the threshold-based emergency stop prediction is also directly affected because both the instantaneous and cumulative errors are not sufficiently accurate.

In the ablation study, both training and validation RMSE losses decreased with increasing epochs, which is typical for all 168 cases. Fig. 11 depicts the loss behavior for the best architecture in Table I.

VI. Conclusion and Future Works

In this study, we developed a novel DL-based method for predicting emergency stops in a robotic rehabilitation training system based on a bi-LSTM model. The effectiveness of the proposed method was validated using real patient gait training data collected over several years. The proposed CE method, which excludes crossing cases, is suitable for detecting abnormal emergency stops in robotic gait rehabilitation systems and is relatively robust when compared to instantaneous error.

We hypothesize that similar problems can occur in any wearable robotic system in which humans wear electromechanical devices on their body parts. Therefore, the proposed emergency prediction method can also be applied to other fields that use wearable robots.

In future work, we will investigate how to avoid emergency stops in real-time after they are anticipated using the proposed method for lower-limb rehabilitation training. Emergency stops can be mitigated by reducing the weight bearing, walking speed, and stride length. However, the effectiveness of these mitigation methods for rehabilitation training has not been validated because of the uncertain causes of emergency stops. Therefore, the true causes of emergency stops should be identified (or classified) in the future. Once a cause is identified, the optimal settings for the training parameters can be derived.

References

[1] K. P. Westlake and C. Patten, "Pilot study of lokomat versus manual-assisted treadmill training for locomotor recovery post-stroke," *J. NeuroEng. Rehabil.*, vol. 6, no. 1, pp. 1–11, Dec. 2009, doi: 10.1186/1743-0003-6-18.

[2] A. Pennycott, D. Wyss, H. Vallery, V. Klamroth-Marganska, and R. Riener, "Towards more effective robotic gait training for stroke rehabilitation: A review," *J. Neuroeng. Rehabil.*, vol. 9, p. 65, 2012, doi: 10.3929/ethz-b-000163684.

[3] J.-H. Park, Y.-I. Shin, J. H. You, and M. S. Park, "Comparative effects of robotic-assisted gait training combined with conventional physical therapy on paretic hip joint stiffness and kinematics between subacute and chronic hemiparetic stroke," *NeuroRehabilitation*, vol. 42, no. 2, pp. 181–190, Mar. 2018, doi: 10.3233/NRE-172234.

[4] I. J. Park *et al.*, "Comparative effects of different assistance force during robot-assisted gait training on locomotor functions in patients with subacute stroke: An assessor-blind, randomized controlled trial," *Amer. J. Phys. Med. Rehabil.*, vol. 98, no. 1, pp. 58–64, Jan. 2019, doi: 10.1097/PHM.0000000000001027.

[5] C. Park *et al.*, "Effects of innovative hip-knee-ankle interlimb coordinated robot training on ambulation, cardiopulmonary function, depression, and fall confidence in acute hemiplegia," *NeuroRehabilitation*, vol. 46, no. 4, pp. 577–587, Jul. 2020, doi: 10.3233/NRE-203086.

[6] J. S. Seo, H. S. Yang, S. Jung, C. S. Kang, S. Jang, and D. H. Kim, "Effect of reducing assistance during robot-assisted gait training on step length asymmetry in patients with hemiplegic stroke: A randomized controlled pilot trial," *Medicine*, vol. 97, no. 33, Aug. 2018, Art. no. e11792, doi: 10.1097/MD.0000000000011792.

[7] J. Hwang, Y. Shin, J.-H. Park, Y. J. Cha, and J. H. You, "Effects of Walkbot gait training on kinematics, kinetics, and clinical gait function in paraplegia and quadriplegia," *NeuroRehabilitation*, vol. 42, no. 4, pp. 481–489, Jun. 2018, doi: 10.3233/NRE-172226.

[8] K. van Kammen, A. M. Boonstra, L. H. V. van der Woude, C. Visscher, H. A. Reinders-Messelink, and R. den Otter, "Lokomat guided gait in hemiparetic stroke patients: The effects of training parameters on muscle activity and temporal symmetry," *Disab. Rehabil.*, vol. 42, no. 21, pp. 2977–2985, 2020, doi: 10.1080/09638288.2019.1579259.

[9] H. E. Yang *et al.*, "Structural and functional improvements due to robot-assisted gait training in the stroke-injured brain," *Neurosci. Lett.*, vol. 637, pp. 114–119, Jan. 2017, doi: 10.1016/j.neulet.2016.11.039.

[10] J. H. Jung, N. G. Lee, J. H. You, and D. C. Lee, "Validity and feasibility of intelligent walkbot system," *Electron. Lett.*, vol. 45, no. 20, pp. 1016–1017, 2009, doi: 10.1049/el.2009.0879.

[11] A. Arami, A. Poulakakis-Daktylidis, Y. F. Tai, and E. Burdet, "Prediction of gait freezing in parkinsonian patients: A binary classification augmented with time series prediction," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 9, pp. 1909–1919, Sep. 2019, doi: 10.1109/TNSRE.2019.2933626.

[12] S. T. Moore *et al.*, "Autonomous identification of freezing of gait in Parkinson's disease from lower-body segmental accelerometry," *J. NeuroEng. Rehabil.*, vol. 10, no. 1, p. 19, 2013, doi: 10.1186/1743-0003-10-19.

[13] L. Palmerini, L. Rocchi, S. Mazilu, E. Gazit, J. M. Hausdorff, and L. Chiari, "Identification of characteristic motor patterns preceding freezing of gait in Parkinson's disease using wearable sensors," *Frontiers Neurol.*, vol. 8, pp. 1–12, Aug. 2017, doi: 10.3389/fneur.2017.00394.

[14] W. Liu, W. Luo, D. Lian, and S. Gao, "Future frame prediction for anomaly detection—A new baseline," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6536–6545, doi: 10.1109/CVPR.2018.00684.

[15] M. Khokhlova, C. Migniot, A. Morozov, O. Sushkova, and A. Dipanda, "Normal and pathological gait classification LSTM model," *Artif. Intell. Med.*, vol. 94, pp. 54–66, Mar. 2019, doi: 10.1016/j.artmed.2018.12.007.

[16] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, nos. 5–6, pp. 602–610, Jul. 2005, doi: 10.1016/j.neunet.2005.06.042.

[17] J. Dicharry, "Kinematics and kinetics of gait: From lab to clinic," *Clinics Sports Med.*, vol. 29, no. 3, pp. 347–364, Jul. 2010, doi: 10.1016/j.csm.2010.03.013.

[18] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecasting*, vol. 22, no. 4, pp. 679–688, Oct. 2006, doi: 10.1016/j.ijforecast.2006.03.001.

[19] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, vol. 1, 2015, pp. 448–456.

[20] R. Khusainov, D. Azzi, I. Achumba, and S. Bersch, "Real-time human ambulation, activity, and physiological monitoring: Taxonomy of issues, techniques, applications, challenges and limitations," *Sensors*, vol. 13, no. 10, pp. 12852–12902, Sep. 2013, doi: 10.3390/s131012852.

[21] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. Conf. Track (ICLR)*, 2015, pp. 1–15.

[22] T. R. Shultz and S. E. Fahlman, *Encyclopedia of Machine Learning and Data Mining*. Boston, MA, USA: Springer, 2017.

[23] M. H. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine," *Clin. Chem.*, vol. 39, no. 4, pp. 561–577, Apr. 1993, doi: 10.1093/clinchem/39.4.561.