ALANet:Autoencoder-LSTM for pain and protective behaviour detection

Xinhui Yuan[1], Marwa Mahmoud[2]
[1] School of Science, ChangChun University of Science and Technology, ChangChun, China
[2] Department of Computer Science and Technology, University of Cambridge, Cambridge, United Kingdom

***Abstract*— Automatic detection of pain and protective behaviour can help chronic pain patients to get proper assistance and helpful treatment with the help of medical professionals. Using the EmoPain datasetwe study how autoencoder-based and attention-based deep learning models can be used to automatically detect pain and protective behavior that is usually associated with it. We propose a deep learning architecture called Autoencoder-LSTM-Attention-Net (ALANet), which can improve the automatic detection of pain and protective behaviors. Through comparative experiments with other machine learning models trained on the EmoPain dataset, we found that by using a combination of autoencoder and attention mechanisms, we can not only improve the recognition performance, but also greatly increase the speed of training the model. In addition, we analyse the effect of extracting temporal information from each body part separately compared to all body parts combined.**

I. INTRODUCTION

Fear of pain and injury in people with chronic pain results in reducing physical activity or using movement strategies (e.g. guarding, stiffness, hesitation, bracing) [1][2][3], collectively called protective behaviors [4]. Such behaviors cause further debilitation and reduced participation in valued activities, e.g. employment or social life [2][5]. Automatic detection of pain and protective behaviour can help those patients get further help from other people. At present, the number of researches on automatic analysis of emotion-influenced movement behavior is very small, as most studies in pain-related situations focus on facial expression of pain or physiological responses to acute/stimulated pain [6][7]. This is partly because of the limited data available for body movements and not enough previous work done in this field compared to facial expression analysis [8][9].

In order to study pain and protective behavior, we used the EmoPain dataset[10]. The dataset contains 26 joint-based whole-body motion capture (MoCap) data and 4 skin surface electromyography (sEMG) data. Previous work on Mocap and sEMG data of the EmoPain dataset mainly used feature engineering methods [10][11] [12], Stacked-LSTM [13], or attention mechanisms [14]. In this paper, we propose an architecture called ALANet, which is a combination of autoencoders, LSTMs, and attention based deep learning architecture. This structure can analyze different actions and extract discriminative features that can best determine when pain signals and protective behavior are exhibited in the data. Then it automatically analyzes when - temporal attention - and what - bodily attention - subsets of the joint-based data contribute most to the detection of pain and protective behaviour[14]. Our contributions can be summarized as:

1. We propose a deep learning architecture using a combination of autoencoders, LSTMs and attention mechanisms to automatically detect and classify pain and protective behaviour.

2.We demonstrate experimentally that our model outperforms the state-of-the-art models while increasing the training speed by reducing the dimensions of the raw data using the autoencoder layers.

3. We experiment with extracting temporal information from the whole body and from each body part separately to show the role of the autoencoder structure in feature extraction and the potential associations between different body parts.

II. RELATED WORKS

One of the first experiments on EmoPain dataset was done by Aung *et al.* [10]. They used a random forest(RF) algorithm to study the MoCap and sEMG data to detect the protective behaviours in some specific exercises. Then, Wang *et al.* [13] conducted a study, using a stacked LSTM architecture to detect protective behavior using a sliding window approach extracted from different movements.

Numerous experiments show that using attention mechanism can greatly improve the performance of automatic human activity recognition (HAR). Zeng *et al.* [15] applied temporal attention to the hidden layer of LSTM, filtered out the irrelevant parts of the data and applied sensor attention to the input layer of LSTM to obtain the information of important sensors. Their works solved two problems in previous research: 1) Some useful information only appears in a short time interval, 2) Some unimportant sensors will mix a lot of noise into the data. Then they also propose continuous attention constraints to further improve performance. Murahari *et al.* [16] used a similar attention mechanism. They added the attention layer to the end of a deepconvlsm [17], without changing any other hyper-parameters. Yao *et al.* [18] added sensor attention to the lower layers of DeepSense framework [19] and added temporal attention to the higher layers of [19], which got better performance than the results achieved by the original DeepSense framework [19]. A more recent study was done by Wang *et al.* [14], which achieved the state-of-the-art results so far. They proposed a network named BANET. This architecture combined LSTM with temporal attention (convolution layer and softmax classifier)

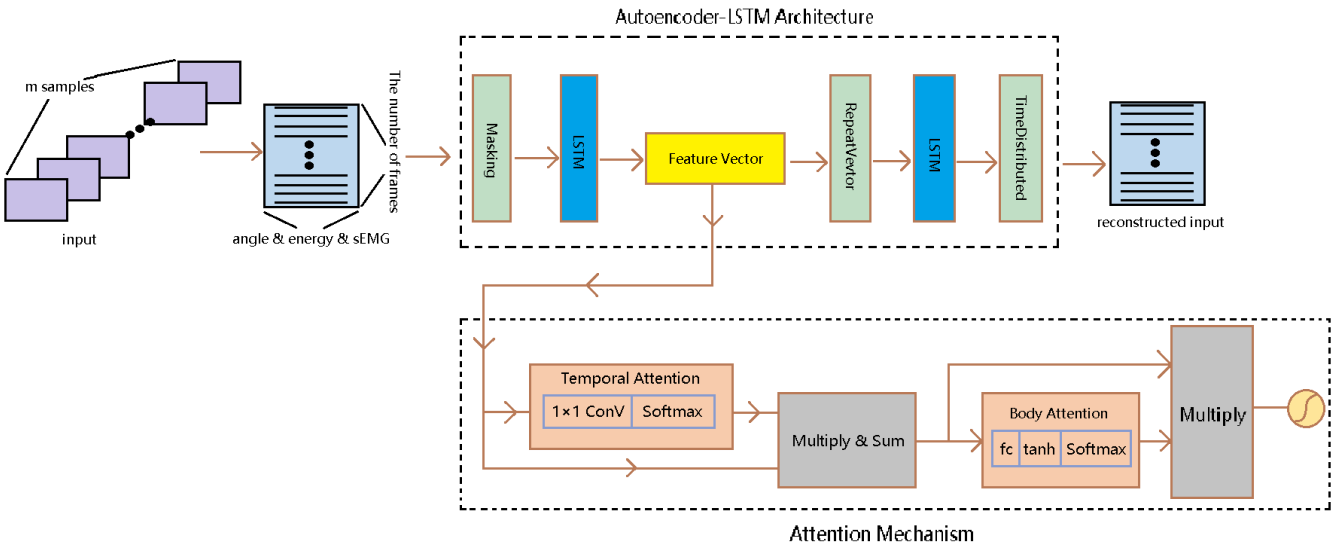

Fig. 1. An overview of the Autoencoder-LSTM-Attention-Net. Each sample contains joint angle, energy and sEMG features.

and bodily attention (full connected layer, tanh activation function and softmax classifier) to learn the features extracted from MoCap and sEMG data. As shown in previous work, using attention mechanism can greatly improve the recognition performance of HAR. Our proposed approach extends the attention mechanism methods by integrating it with Autoencoder and LSTM architecture in order to extract better feature representation instead of applying attention mechanisms directly on raw data.

III. METHODLOGY

In this section, we present our proposed deep learning architecture that we use for pain detection and protective behaviour detection. An overview of the proposed architecture is shown in Fig. 1, where we show how we combine Autoencoder LSTMs with an attention network.

A. Autoencoder LSTM

The first step in our network is an Autoencoder LSTM network. Autoencoders [20] have been used in many previous works to encode raw data by reducing the dimensionality of the data. We combine autoencoder structure with LSTM[21] neural network, which gives the autoencoder structure the ability to process and encode temporal information. This is useful to produce a concise temporal representation of the raw features. Our autoencoder network architecture has five layers: Masking, LSTM, RepeatVector, LSTM and TimeDistributed layer. The input is body movement features: angles, energy and sEMG. The features extracted by the first LSTM layer are reconstructed to produce an output as close as possible to the original input data. Then the features extracted from the autoencoder LSTM are passed to the attention layers.

B. Attention Mechanism

As mentioned in section II, attention mechanisms have been used successfully in body movement related tasks. After the temporal body features are encoded using the Autoencoder LSTM, it is passed to the attention layers. Our attention layers are based on the work described in [14]. Using attention layers allows the network to pay more attention to the most salient features (body attention) and frames (temporal attention). Temporal attention layer is a convolution layer with softmax activation. Body attention layers contain one full connected layer with tanh activation and one fully connected layer with softmax activation. After the attention layers, the last part in our architecture is one fully connected layer with softmax activation to produce the final output.

IV. DATA PREPARATION

A. Data Segmentation

To evaluate our model, we use EmoPain dataset [10] as part of Emopain2020 challenge[22]. The dataset is collected from 14 patients with chronic pain and 9 healthy participants. The training set contains 10 chronic pain participants and 6 healthy participants. The validation set contains 4 chronic pain participants and 3 healthy participants. The features extracted from each sample video are the angles of the 13 whole-body joints, the energy of each angle(The energy is based on the square of each angular velocity. The positions of the 13 joints in the whole body are shown in the Fig. 2 [22]) and the sEMG data obtained from the 4 skin electromyography sensors (shown in Fig. 3 [22]). We tested our proposed model on task two and task three sub-challenges [22]. Task two[23] is about pain detection. The training set contains 398 samples containing 226 samples labeled with 'healthy', 92 samples labeled with 'low-level pain' and 80 samples labeled with 'high-level pain'. The testing set contains 416

S/N	Angle	Incident joint	Endpoint joints	Anatomical positions
1.	Left Full-body Flexion	lower spine	crown, left ankle	26-1-4
2.	Right Full-body Flexion	lower spine	crown, right ankle	26-1-9
3.	Left Inner-body Flexion	lower spine	mid spine, left knee	12-1-3
4.	Right Inner-body Flexion	lower spine	mid spine, right knee	12-1-8
5.	Left Knee Angle	left knee	left hip, left ankle	2-3-4
6.	Right Knee Angle	right knee	right hip, right ankle	7-8-9
7.	Left Elbow Angle	left elbow	left upper arm, left hand	15-16-17
8.	Right Elbow Angle	right elbow	right upper arm, right hand	20-21-22
9.	Left Shoulder Angle	left shoulder	neck, left upper arm	24-14-15
10	Right Shoulder Angle	right shoulder	neck, right upper arm	24-19-20
11	Left Lateral Bend	left shoulder	left elbow, left hip	16-14-2
12	Right Lateral Bend	right shoulder	right elbow, right hip	21-19-7
13	Neck Angle	neck	mid spine	12-24-26

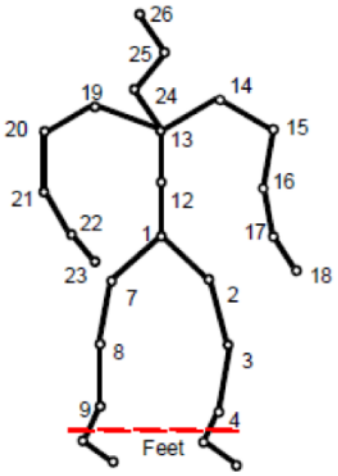

Fig. 2. Joint angle illustration as explained in [22]

samples containing 313 samples labeled with 'healthy', 89 samples labeled with 'low-level pain' and 14 samples labeled with 'high-level pain'. Task three [24] is about protective behaviour detection. As described in the challenge, input samples are defined using a sliding window approach [13] with an overlap of 75%. Since the labels for data are provided for every frame, in order to generate one label for every sliding window, we use a majority-voting approach to assign the most frequent label to the whole window segment. This produces 6440 samples in the training set containing 5334 non-protective behaviour and 1106 protective behaviour, and 2833 samples in the validation set containing 2651 non-protective behaviour and 182 protective behaviour.

Fig. 3. sEMG data from 4 sensors on the back as described in [22]

B. Data Augmentation

As shown in the previous section, the data is not very well balanced. For example, the protective behaviour task has 182 samples compared to 2651 non-protective behaviour in the validation set. To reduce the impact of small dataset and unbalanced labels, we apply two data augmentation methods on the training set of both tasks, based on the methods described in [13][14]. The first method creates new instances by adding normalized Gaussian noise to the original data with 3 different standard deviations: 0.05, 0.1 and 0.15 [14]. The second approach creates new instances by randomly setting the data of some frames and some body parts to 0 with selection probability of 0.05, 0.1 and 0.15 [13]. These methods were chosen because they simulate missing data and extend the size of the dataset without the risk of overfitting. After augmenting the data in the training set, we finally got 1882 samples containing 678 healthy, 644 low-level pain and 560 high-level pain in the pain detection task and 18410 samples, containing 10668 non-protective behaviour and 7742 protective behaviour in the protective behaviour detection task.

V. Experimented Evaluation

A. Implementation Details

We trained our ALANet using Keras with a TensorFlow back-end. For the Autoencoder-LSTM part, we used Adam optimizer [25] to update the weights. The learning rate and batch size were set to the default sizes of Keras. We used MSE function to be the loss function. For the attention layers part, we used the same implementation of the attention mechanism used in [14]. The 1 * 1 convolution layer and softmax layers are used as temporal attention, and the fully connected layer, tanh activation function, and softmax activation function are used as body attention. The temporal feature output from the first LSTM layer of the Autoencoder-LSTM part is used as the input to the attention part. As for the structure of the attention mechanism, we only change the number of hidden units in the fully connected layer to correspond to the input data, without changing other parameter settings. We also used the Adam optimizer [25] with learning rate=0.003 and batch size=40. Two methods were used to evaluate our proposed approach. One method is hold-out validation, where we use the training set to train the model and the validation set to evaluate it. The second method is leave-one-subject-out cross-validation (LOSOCV).

TABLE I

RESULTS FOR PROTECTIVE BEHAVIOUR DETECTION TASK USING HOLD-OUT VALIDATION

Model	ACC	F1-Score		Training Time
Stacked-LSTM [26]	0.4636	0	–	–
		1	–	
		Average	0.48	
BANet [14]	0.77	0	0.86	$>$ 10 hours
		1	0.28	
		Average	0.57	
ALANet	**0.85**	0	0.91	$<$ **30 minutes**
		1	0.26	
		Average	**0.59**	

B. Protective Behaviour Detection Task

In protective behaviour detection task, we use hold-out validation method and compare with the challenge baseline model [26] and BANet [14]. The results are shown in Table I. Our model(ALANet) achieves an accuracy of 0.851, and mean F1-score of 0.587, which are significantly better than the baseline results [26](accuracy of 0.4636, mean F1-score of 0.4811). It also achieves comparable results to BANet [14](accuracy of 0.77, mean F1-score of 0.57). In terms of speed of training the model, our model only takes less than 30 minutes, while BANet takes more than 10 hours. Compared to BANet, although the increase in performance of our model is not significantly high, the training speed improves a lot. This is due to the use of the autoencoder LSTM step, which decreases the size of the raw data significantly before applying the attention layers.

Moreover, we study the effect of encoding the temporal features of every body part separately compared to the whole feature set combined. We do that by changing the architecture presented in section III and use the autoencoder on every feature - body part - separately. As shown in table II, encoding every body part separately performed worse than encoding all body parts together. This may be because there is usually an association between various body parts especially during specific actions or movements. Therefore, encoding the temporal information for all body parts together takes care of this association between different body parts.

C. Pain Detection Task

Pain detection task results are shown in Table III. Here we use two evaluation methods: hold-out validation and leave-one-subject-out cross-validation (LOSOCV). As shown in table III, when using hold-out validation, our model achieves comparable results to KNN and SVM mentioned in [26]. But in accuracy, our model achieves the best results(accuracy of 0.66), which outperforms the other two methods. When using LOSOCV, KNN achieves an accuracy of 0.37 and mean F1-score of 0.34, while SVM achieves an accuracy of 0.44 and mean F1-score of 0.41. Our model achieves the best results (accuracy of 0.56 and mean F1-score of 0.41)

TABLE II

BANET VS. ALANET RESULTS WHEN EXTRACTING TEMPORAL FEATURES FROM EVERY BODY PART SEPARATELY USING HOLD-OUT VALIDATION

Model	Acc	F1-Score	
BANet [14]	0.54	0	0.69
		1	0.09
		Average	0.39
ALANet	0.78	0	0.87
		1	0.18
		Average	0.53

TABLE III

RESULTS FOR PAIN DETECTION TASK USING LOSOCV AND HOLD-OUT VALIDATION

Model	Evaluate method	Acc	F1-Score	
KNN [26]	LOSOCV	0.37	Average	0.34
	Hold-out	0.35	0	0.39
			1	0.09
			2	0.44
			Average	0.31
SVM(Sigmoid/ Gaussian kernels) [26]	LOSOCV	0.44	Average	0.41
	Hold-out	0.07	0	0
			1	0.14
			2	0
			Average	0.34
ALANet	LOSOCV	**0.526**	Average	**0.426**
	Hold-out	**0.4561**	0	0.403
			1	0
			2	0
			Average	0.134

VI. CONCLUSIONS

This paper proposes a novel architecture ALANet that uses autoencoder LSTMs with attention mechanisms to automatically detect pain levels and protective behaviour in the movement of patients suffering from chronic pain. In protective behaviour detection task, when comparing our proposed approach with the baseline model [26] and state-of-the-art models, our approach manages to achieve better results with the least training time. The results show that the Autoencoder-LSTM structure is more helpful than single LSTM layers in extracting tamporal information. As for training time, our model achieves the fastest training time (less than 30 minutes). The Autoencoder-LSTM structure only focuses on the most discriminative temporal information. In pain detection task, the accuracy of our model outperformed the baseline model [26] despite the fact that the mean F1-score did not improve much. We think that this is because of the limited number of samples for specific categories and the unbalanced dataset. For future work, we would like to evaluate our model on bigger datasets and also on different tasks other than pain detection, such as behaviour modelling for psychological distress.

REFERENCES

[1] Johan WS Vlaeyen and Steven J Linton. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain*, 85(3):317–332, 2000.

[2] Johan WS Vlaeyen, Stephen Morley, and Geert Crombez. The experimental analysis of the interruptive, interfering, and identity-distorting effects of chronic pain. *Behaviour research and therapy*, 86:23–34, 2016.

[3] Temitayo Olugbade, Nadia Bianchi-Berthouze, and Amanda C de C Williams. The relationship between guarding, pain, and emotion. *PAIN Report*, 4(4), 2019.

[4] Francis J Keefe and Andrew R Block. Development of an observation method for assessing pain behavior in chronic low back pain patients. *Behavior Therapy*, 1982.

[5] Harald Breivik, Beverly Collett, Vittorio Ventafridda, Rob Cohen, and Derek Gallacher. Survey of chronic pain in europe: prevalence, impact on daily life, and treatment. *European journal of pain*, 10(4):287–287, 2006.

[6] Beat Fasel and Juergen Luettin. Automatic facial expression analysis: a survey. *Pattern recognition*, 36(1):259–275, 2003.

[7] Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE transactions on pattern analysis and machine intelligence*, 31(1):39–58, 2008.

[8] Michael JL Sullivan, Pascal Thibault, André Savard, Richard Catchlove, John Kozey, and William D Stanish. The influence of communication goals and physical demands on different dimensions of pain behavior. *Pain*, 125(3):270–277, 2006.

[9] Beatrice De Gelder. Why bodies? twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3475–3484, 2009.

[10] Min SH Aung, Sebastian Kaltwang, Bernardino Romera-Paredes, Brais Martinez, Aneesha Singh, Matteo Cella, Michel Valstar, Hongying Meng, Andrew Kemp, Moshen Shafizadeh, et al. The automatic detection of chronic pain-related expression: requirements, challenges and the multimodal emopain dataset. *IEEE transactions on affective computing*, 7(4):435–451, 2015.

[11] Temitayo A Olugbade, MS Aung, Nadia Bianchi-Berthouze, Nicolai Marquardt, and Amanda C Williams. Bi-modal detection of painful reaching for chronic pain rehabilitation systems. In *Proceedings of the 16th International Conference on Multimodal Interaction*, pages 455–458. ACM, 2014.

[12] Temitayo A Olugbade, Nadia Bianchi-Berthouze, Nicolai Marquardt, and Amanda C Williams. Pain level recognition using kinematics and muscle activity for physical rehabilitation in chronic pain. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 243–249. IEEE, 2015.

[13] Chongyang Wang, Temitayo A Olugbade, Akhil Mathur, Amanda C De C Williams, Nicholas D Lane, and Nadia Bianchi-Berthouze. Recurrent network based automatic detection of chronic pain protective behavior using mocap and semg data. In *Proceedings of the 23rd International Symposium on Wearable Computers*, pages 225–230. ACM, 2019.

[14] Chongyang Wang, Min Peng, Temitayo A Olugbade, Nicholas D Lane, Amanda C de C Williams, and Nadia Bianchi-Berthouze. Learning temporal and bodily attention in protective movement behavior detection. pages 324–330, 2019.

[15] Ming Zeng, Haoxiang Gao, Tong Yu, Ole J Mengshoel, Helge Langseth, Ian Lane, and Xiaobing Liu. Understanding and improving recurrent networks for human activity recognition by continuous attention. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers*, pages 56–63. ACM, 2018.

[16] Vishvak S Murahari and Thomas Plötz. On attention models for human activity recognition. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers*, pages 100–103. ACM, 2018.

[17] Francisco Ordóñez and Daniel Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1):115, 2016.

[18] Shuochao Yao, Yiran Zhao, Shaohan Hu, and Tarek Abdelzaher. Qualitydeepsense: Quality-aware deep learning framework for internet of things applications with sensor-temporal attention. In *Proceedings of the 2nd International Workshop on Embedded and Mobile Deep Learning*, pages 42–47. ACM, 2018.

[19] Shuochao Yao, Shaohan Hu, Yiran Zhao, Aston Zhang, and Tarek Abdelzaher. Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In *Proceedings of the 26th International Conference on World Wide Web*, pages 351–360. International World Wide Web Conferences Steering Committee, 2017.

[20] Andrew Ng. Sparse autoencoder. *CS294A Lecture notes*, 72(2011):1–19, 2011.

[21] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[22] https://mvrjustid.github.io/emopainchallenge2020/.

[23] Temitayo A Olugbade, Aneesha Singh, Nadia Bianchi-Berthouze, Nicolai Marquardt, Min SH Aung, and Amanda C De C Williams. How can affect be detected and represented in technological support for physical rehabilitation? *ACM Transactions on Computer-Human Interaction (TOCHI)*, 26(1):1–29, 2019.

[24] MS Hane Aung, Nadia Bianchi-Berthouze, Paul Watson, and AC de C Williams. Automatic recognition of fear-avoidance behavior in chronic pain physical rehabilitation. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, pages 158–161. ICST (Institute for Computer Sciences, Social-Informatics and ..., 2014.

[25] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[26] Joy Egede, Temitayo Olugbade, Chongyang Wang, Siyang Song, Nadia Berthouze, Michel Valstar, Amanda Williams, Hongyin Meng, Min Aung, and Nicholas Lane. Emopain challenge 2020: Multimodal pain evaluation from facial and bodily expressions. *arXiv preprint arXiv:2001.07739*, 2020.