PRELIMINARY RESULTS ON THE GENERATION OF ARTIFICIAL HANDWRITING DATA USING A DECOMPOSITION-RECOMBINATION STRATEGY

José Fernando Adrán Otero [1,2] *Oscar Soláns Caballer* [1,2] *Pere Marti-Puig* [2] *Zhe Sun* [3]
Toshihisa Tanaka [4,5] *Jordi Solé-Casals* [2,6*]

[1]Faculty of Computer Science, Multimedia and Telecommunications, Open University of Catalonia, 08080 Barcelona, Spain

[2]Data and Signals Processing Research Group, University of Vic-Central University of Catalonia, 08500 Barcelona, Spain

[3]Computational Engineering Applications Unit, Head Office for Information Systems and Cybersecurity, RIKEN, 351-0198 Saitama, Japan

[4]Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology, Japan

[5]Tensor Learning Team, RIKEN Center for Advanced Intelligence Project, Japan

[6]Department of Psychiatry, University of Cambridge, Cambridge CB2 3EB, U.K.

ABSTRACT

Deep learning techniques are able to extract the characteristics of temporal signals to study their patterns and diagnose diseases such as essential tremor. However, these techniques require a large amount of data to train the neural network and achieve good results, and the more data the network has, the more accurate the final model implemented will be. This work proposes the use of data augmentation techniques to improve the accuracy of a Long short-term memory system in the diagnosis of essential tremor. For this purpose, the Empirical Modal Decomposition method will be used to decompose the original temporal signals collected from control subjects and patients with essential tremor. The time series obtained from the decomposition, covering different frequency ranges, will be randomly shuffled and combined to generate new artificial samples for each group. Then, both the generated artificial samples and part of the real samples will be used to train the LSTM network, and the remaining original samples will be used to test the model. Experimental results demonstrate the capability of the proposed method, increasing the classifier accuracy from 83.2% to almost 93% when artificial samples are used.

Index Terms— Deep Learning, LSTM, Data Augmentation, Empirical Mode Decomposition, Essential Tremor

1. INTRODUCTION

Essential tremor (ET) is a disorder of the nervous system causing involuntary, rhythmic movements that can appear from the age of 40 years but usually appears from the age of 65 years onward [1]. The tremor usually appears in the hands, and especially when the subject performs simple everyday tasks, such as eating or tying shoelaces. Although it is not usually a serious condition, it can become so if severe, and can be confused with Parkinson's disease. In this work we will use a deep neural network based on LSTM [2, 3] to perform this diagnosis. Given that the number of parameters of the deep model is very high, hundreds or thousands of samples are needed to train it. Overfitting occurs when a system becomes too specialised with the training data and is unable to successfully classify new test data. With small datasets, overfitting is a problem to be taken into account. To reduce or avoid this problem, the input dataset should be larger, which can be solved using data augmentation.

Data augmentation refers to the process of synthesising new data from the existing available data. One possible option is to apply various transformations to real data to synthesise new data [4]. In the handwriting domain, authors in [5] proposes a method designed to learn proper and efficient data augmentation which is more effective and specific for training a robust recognizer, and in [6] an Adversarial Feature Deformation Module (AFDM) that learns ways to elastically warp extracted features in a scalable manner was proposed. In this work we will use a method based on a decomposition-recombination technique to generate artificial samples. This method was firstly proposed in [7] in an EEG framework, and here will be used in handwriting signals for the first time.

2. METHODS

The different steps to generate artificial random samples for training the model are explained below: First, the working scenario is presented and the samples are divided in two groups: ET (essential tremor patients) and CT (control subjects). Then, using the Empirical Mode Decomposition (EMD) [8], the so-called Intrinsic Mode Functions (IMFs) of all the samples in the dataset are obtained. Finally, the new

This work is supported in part by the University of Vic-Central University of Catalonia under the grant R0947, and partially based upon work from COST Action CA18106, supported by COST (European Cooperation in Science and Technology)

artificial samples are generated using only the IMFs from the samples of the training part.

2.1. Scenario

The database used in this work is known as BIODARW. This data was firstly presented in [9] and has been used in previous works [9, 10, 11] to analyze and classify samples (drawings) from ET patients and controls. The participants in the study performed, among other exercises, the test of drawing the Archimedes' Spiral [12] with both hands, using a digitizing tablet to acquire the samples. In order to compare our results with the previous mentioned [9, 10, 11], we will use the same subset, BIODARWO, which consists of 51 samples: 24 samples of the ET group and 27 samples of the control group. A more detailed description of the database can be found at [9, 10, 11].

The data available in this database was obtained from a Wacom 4 digitizing tablet at a sampling frequency of 100 Hz. Among all the possible features collected by the tablet (X coordinate, Y coordinate, time stamp, azimuth angle, angular angle and pressure), we will only use the X and Y coordinates, as it was done in [11]. The reasons are twofold: (i) any digital tablet can obtain these parameters, but maybe not the others (angles, pressure, etc.); (ii) X and Y coordinates carry enough information to diagnose essential tremor, as demonstrated in [11].

2.2. Empirical Mode Decomposition

EMD is based on a method that allows decomposing the different time signals into a finite number of the so called Intrinsic Mode Functions (IMFs). Each IMF represents a non-linear oscillation of the original signal, which can be reconstructed by summing all the IMFs of the signal and its residual (the IMF corresponding to the linear trend). All the IMFs have to fulfill the following two conditions:

1. The number of maximums has to be the same as the number of zero crossings, or at least they have to differ by only one.
2. For any sample, the mean value between the envelope of the local maxima and that of the local minima must be zero.

The algorithm of this decomposition method consists in rewriting a real-valued signal $x(t)$ as: $x(t) = \sum_n x_n(t) + r(t)$ where $x_n(t)$ are the so-called IMFs and $r(t)$ is the residue signal.

The iterative process to obtain them is described below:

1. Define $s(t) = r_{n-1}(t)$. Initialize $n = 1$, $r_0(t) = x(t)$.
2. Extract the n-th IMF as follows:
   (a) Identify all local maxima and minima of $s(t)$.
   (b) Interpolate between the maximum (minimum) to obtain the upper envelope (lower envelope).
   (c) Obtain the local mean $m(t)$ by taking the mean of both envelopes.
   (d) Obtain an IMF candidate by subtracting the local mean $m(t)$ to the signal $s(t)$: $h(t) = s(t) - m(t)$.
   (e) If $h(t)$ does not comply with the two conditions to become an IMF then go to step 2 with $s(t) = h(t)$
3. If $h(t)$ satisfy the IMF conditions, then: $x_n(t) = h(t)$ and $r_n(t) = r_{n-1}(t) - x_n(t)$
4. If $r_n(t)$ is a monotone function, or does not have enough extreme points to calculate the upper and lower envelopes, then $x_n(t)$ is the last IMF function of $x(t)$ and the decomposition ends.
5. Otherwise, set $s(t) = r_n(t)$ and iterate from step 2 to obtain the next IMF.

In order to simultaneously decompose X and Y coordinates for each sample and for all the subjects, the mEMD [13] algorithm was used. In this way, the maximum and minimum envelope will be the same for all subjects and coordinates, and the number of IMFs will be the same for all of them. The decision to use mEMD instead of EMD was made to overcome the problem of having a different number of IMFs depending on the subject, due to the intrinsic data-driven characteristics of the EMD algorithm. This effect resulted in a desynchronisation of weights which meant that the artificial samples did not work properly, clearly deteriorating the accuracy of the classification system. One example of the decomposition is presented in Figure 1, which corresponds to a control subject.

2.3. New artificial samples

The method to create artificial samples is based on previous works in which this method was first presented and used on EEG data for decreasing the calibration step in a BCI application [14], and then used in [7] as a data augmentation technique to train a deep learning classifier, also on EEG data. To our knowledge, this is the first time this method has been used on handwriting data and in a 2D scenario, creating artificial drawings of the Archimedes' Spiral.

Once the IMFs for each subject and coordinate (X and Y) have been obtained, artificial samples can be generated. We generate one artificial sample, of one of the two groups, by randomly selecting the IMF1 from one training subject of the corresponding class, the IMF2 from another subject, the IMF3 from another one, etc., up to the maximum number of IMFs obtained in the mEMD decoposition. Finally the new artificial sample is obtained by summing the selected IMFs. Because of the random selection of the origin of the IMFs, we can repeat this process to obtain the desired amount of

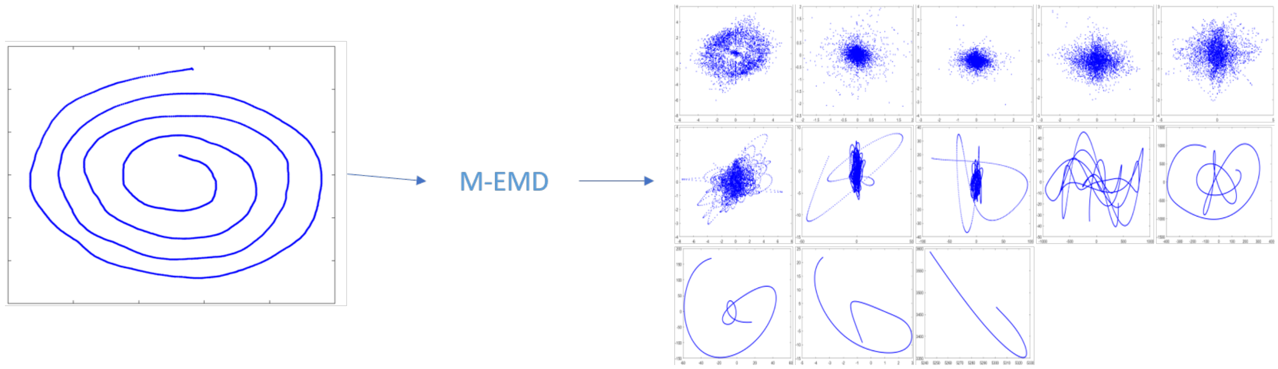

Fig. 1. Extraction of IMFs through mEMD. On the left, an original representation of an Archimedes' Spiral performed by a control subject. On the right, the 13 IMF's obtained by the mEMD algorithm.

new artificial samples. One visual example of the process is shown in Figure 2. Because we use information of some subjects (their IMFs) to create the artificial samples, these real subjects will be part of the training dataset, together with the artificial ones. The remaining (non-used) subjects will conform the test dataset.

2.4. LSTM training

Once the artificial samples have been generated, the next step is to train the model. In this work, an LSTM network with 5 biLSTM layers has been designed [15], and 2 regularization methods have been used: L2 and Dropout [16, 17]. To evaluate the results, the accuracy (ACC), sensitivity (SEN) and specificity (SPE) will be used, calculated as follows:

$$AC = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$SE = \frac{TP}{TP+FN} \tag{2}$$

$$SP = \frac{TN}{TN+FP} \tag{3}$$

where TP, TN, FP, and FN are, respectively, the true positive, true negative, false positive and false negative values of the confusion matrix. The model was optimized using the well known adaptive moment estimation algorithm (Adam), that usually works properly with time series signals. It was trained with 25 epochs on a mini-batch size of 48 samples per epoch, shuffling on every epoch the order of the samples to get more variability to the training model. The learning rate was established at 0.001 as initial rate, with a learn drop factor of 0.9 each 10 epochs. Finally, for the regularization parameters, L2 Regularization value was established at $1 \cdot 10^{-4}$, and the Dropout factor between each layer at 0.2.

3. EXPERIMENTAL RESULTS

To check if the proposed data augmentation method is adequate for handwriting signals, we train the LSTM classification model using 0, 50 or 100 artificial samples per group, plus the real training data. Half of the original subjects per group will be used (50% of the original available data). Hence, the case with less data will be when only using real data (i.e., 0 artificial samples per group), while the case with more training data will be when using the 50% of the real data plus 100 artificial samples per group. Since the results obtained when training a model fluctuate if it is executed several times and for different input data, each experiment will be repeated 10 times, with a random selection of subjects used for the training step. In each case, the artificial samples will be generated using the IMFs of the selected original subjects. Table 1 shows the experimental mean results over the 10 repetitions for the accuracy, sensitivity and specificity, depending on the amount of artificial data used to train the classifier.

Table 1. Accuracy, Sensitivity and Specificity of the testing diagnosis with respect to the amount of artificial samples added (AS)

		AC	SE	SP
AS	0	83.2%	68.2%	**95.0%**
	50	88,0%	85,4%	90,0%
	100	**92,4%**	**93,6%**	91,4%

4. DISCUSSION

In this work we present preliminary results of the effect of the decomposition-recombination data augmentation method, based on EMD. To evaluate its effect, we calculated the mean results in terms of accuracy sensitivity and specificity of 10 experiments in different scenarios, in terms of the number of artificial samples used in each group. The numerical values are presented in Table 1. As can be seen, artificial samples always allow for increased accuracy when adding artificial sam-

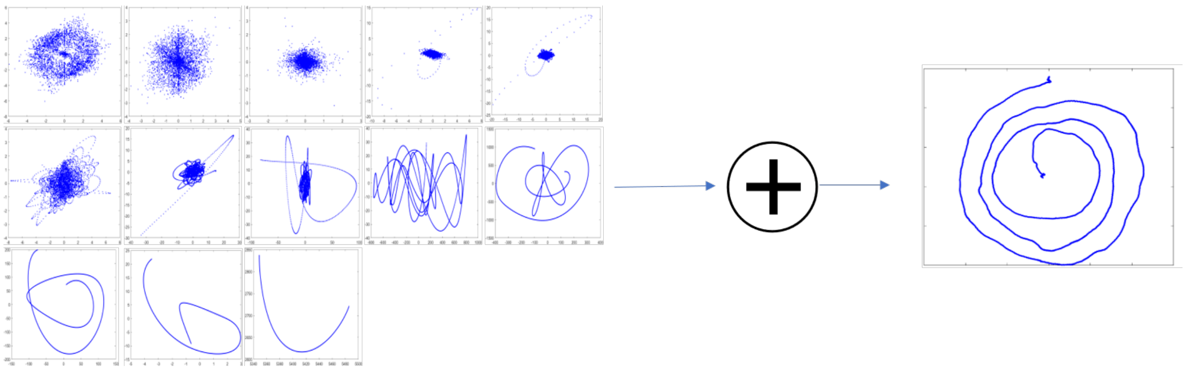

Fig. 2. New sample generation using different IMFs of control random subjects. On the left, a set of IMFs of each category (from IMF1 to IMF13) randomly selected from all the IMFs of the training set. On the right, the visual result of the Archimedes' Spiral obtained after summing all the IMFs.

ples. Taking the case of 0 artificial samples as the baseline, we can observe an increase of about 5% when adding 50 samples and of almost 10% when adding 100 samples, which corresponds to the best case. This is a noticeable improvement compared to the baseline. Sensitivity is also improved in a similar way, from 68.2% to 85.4% first (50 artificial samples), and to 93.6% when adding 100 artificial samples. This is an important result in this application, because in a medical scenario it's of paramount importance to design a system capable of identifying the true positives as much as possible. Finally, the specificity slightly decreases from 95% to 90% first and finally to 91.4%. Even if there is a decrease, the specificity value when using 100 artificial samples is still high, above 90%, which indicates a very good capability of the model to identify true negatives. Overall, the best case is obtained when using 100 artificial samples, with 50% of the original samples for training, and represents an interesting improvement compared to the case without artificial samples, not only in terms of an increasing accuracy, but also being able to balance the sensitivity and the specificity of the model. This is important because it indicates that EMD-based artificial sample generation is useful when we have small data sets, where only a few samples are available to build the model. In the experiments, a total of 25 original samples were used for training (12 from the ET group, 13 from the CT group), to which the artificial samples generated per group were added. The new artificial samples contributed positively to training the LSTM model and controlling the over-fitting of the system. Comparing the best result obtained with the strategy proposed in this paper with those already published in other papers, we can see that the method proposed here is powerful because only using the points of the X and Y coordinates it achieves 92.40% accuracy, which is better than the 91% obtained in [9] using 84 linear and non-linear features extracted from all the available variables generated by the digitizer tablet; and similar to the value obtained in [10] using 77 linear and non-linear features. However, we are below the 97.96% obtained in [11], which obtained this result after selecting the best 5 features over a set of 35 pre-computed linear and non-linear features (in time and frequency domain) and combining two different strategies (residual method and radius method). We can see that this was a complex system based on the pre-calculation of a large set of features, in contrast to the proposed method using LSTM, where no features are calculated beforehand.

5. CONCLUSION

This study presents a deep learning LSTM network for the diagnosis of patients with essential tremor. Since deep neural networks need a large amount of data to converge, a data augmentation method based on a decomposition-recombination strategy using EMD is implemented. The artificial samples avoid over-fitting in the training stage of the LSTM network and increase the classification results in the test stage. Moreover, since the acquisition of the original data requires specific equipment, the study has been carried out using only the X and Y coordinate points, with the idea that it can be easily replicated by acquiring the data using any digitising tablet. The results obtained show a significant improvement in accuracy of almost 10% compared to the case without artificial samples. In addition, the proposed method also increases sensitivity and maintains high sensitivity, resulting in a very well-balanced system with a near-diagonal confusion matrix. In future work we will investigate the effect of increasing the amount of artificial data generated, to test the limits of this method. In addition, we will also investigate faster methods to decompose the samples with EMD, as the mEMD method is very time-consuming. We also plan to compare our method of handwriting data augmentation with other methods based on different approaches. Finally, other configurations of recursive neural networks will be explored with the aim of determining the best possible model for diagnosing essential tremor using only the X and Y coordinates.

References

[1] Elan D Louis and Jean Paul G Vonsattel, "The emerging neuropathology of essential tremor," *Movement Disorders*, vol. 23, no. 2, pp. 174–182, 2008.

[2] Felix A Gers, Jürgen Schmidhuber, and Fred Cummins, "Learning to forget: Continual prediction with lstm," *Neural computation*, vol. 12, no. 10, pp. 2451–2471, 2000.

[3] Sima Siami-Namini, Neda Tavakoli, and Akbar Siami Namin, "The performance of lstm and bilstm in forecasting time series," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 3285–3292.

[4] David A Van Dyk and Xiao-Li Meng, "The art of data augmentation," *Journal of Computational and Graphical Statistics*, vol. 10, no. 1, pp. 1–50, 2001.

[5] Canjie Luo, Yuanzhi Zhu, Lianwen Jin, and Yongpan Wang, "Learn to augment: Joint data augmentation and network optimization for text recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 13746–13755.

[6] Ayan Kumar Bhunia, Abhirup Das, Ankan Kumar Bhunia, Perla Sai Raj Kishore, and Partha Pratim Roy, "Handwriting recognition in low-resource scripts using adversarial learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4767–4776.

[7] Josep Dinarès-Ferran, Rupert Ortner, Christoph Guger, and Jordi Solé-Casals, "A new method to generate artificial frames using the empirical mode decomposition for an eeg-based motor imagery bci," *Frontiers in neuroscience*, vol. 12, pp. 308, 2018.

[8] Gabriel Rilling, Patrick Flandrin, Paulo Goncalves, et al., "On empirical mode decomposition and its algorithms," in *IEEE-EURASIP workshop on nonlinear signal and image processing*. Citeseer, 2003, vol. 3, pp. 8–11.

[9] Karmele López-de Ipiña, Jordi Solé-Casals, Marcos Faundez-Zanuy, Pilar M Calvo, Enric Sesa, Unai Martinez de Lizarduy, Patricia De La Riva, Jose F Marti-Masso, Blanca Beitia, and Alberto Bergareche, "Selection of entropy based features for automatic analysis of essential tremor," *Entropy*, vol. 18, no. 5, pp. 184, 2016.

[10] Karmele Lopez-de Ipina, Jordi Solé-Casals, Marcos Faúndez-Zanuy, Pilar M Calvo, Enric Sesa, Josep Roure, Unai Martinez-de Lizarduy, Blanca Beitia, Elsa Fernández, Jon Iradi, et al., "Automatic analysis of archimedes' spiral for characterization of genetic essential tremor based on shannon's entropy and fractal dimension," *Entropy*, vol. 20, no. 7, pp. 531, 2018.

[11] Jordi Solé-Casals, Iker Anchustegui-Echearte, Pere Marti-Puig, Pilar M Calvo, Alberto Bergareche, José Ignacio Sánchez-Méndez, and Karmele Lopez-de Ipina, "Discrete cosine transform for the analysis of essential tremor," *Frontiers in physiology*, vol. 9, pp. 1947, 2019.

[12] Seth L Pullman, "Spiral analysis: a new technique for measuring tremor with a digitizing tablet," *Movement disorders*, vol. 13, no. S3, pp. 85–89, 1998.

[13] Naveed ur Rehman, Cheolsoo Park, Norden E Huang, and Danilo P Mandic, "Emd via memd: multivariate noise-aided computation of standard emd," *Advances in adaptive data analysis*, vol. 5, no. 02, pp. 1350007, 2013.

[14] Zhiwen Zhang, Feng Duan, Jordi Sole-Casals, Josep Dinares-Ferran, Andrzej Cichocki, Zhenglu Yang, and Zhe Sun, "A novel deep learning approach with data augmentation to classify motor imagery signals," *IEEE Access*, vol. 7, pp. 15945–15954, 2019.

[15] Luke Taylor and Geoff Nitschke, "Improving deep learning with generic data augmentation," in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2018, pp. 1542–1547.

[16] Ekachai Phaisangittisagul, "An analysis of the regularization between l2 and dropout in single hidden layer neural network," in *2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS)*. IEEE, 2016, pp. 174–179.

[17] Yarin Gal and Zoubin Ghahramani, "A theoretically grounded application of dropout in recurrent neural networks," *Advances in neural information processing systems*, vol. 29, pp. 1019–1027, 2016.