Synthetic Biological Signals Machine-generated by GPT-2 improve the Classification of EEG and EMG through Data Augmentation

Jordan J. Bird[1], Michael Pritchard[1], Antonio Fratini[2], Anikó Ekárt[3], Diego R. Faria[1]

Abstract—Synthetic data augmentation is of paramount importance for machine learning classification, particularly for biological data, which tend to be high dimensional and with a scarcity of training samples. The applications of robotic control and augmentation in disabled and able-bodied subjects still rely mainly on subject-specific analyses. Those can rarely be generalised to the whole population and appear to over complicate simple action recognition such as grasp and release (standard actions in robotic prosthetics and manipulators). We show for the first time that multiple GPT-2 models can machine-generate synthetic biological signals (EMG and EEG) and improve real data classification. Models trained solely on GPT-2 generated EEG data can classify a real EEG dataset at 74.71% accuracy and models trained on GPT-2 EMG data can classify real EMG data at 78.24% accuracy. Synthetic and calibration data are then introduced within each cross validation fold when benchmarking EEG and EMG models. Results show algorithms are improved when either or both additional data are used. A Random Forest achieves a mean 95.81% (1.46) classification accuracy of EEG data, which increases to 96.69% (1.12) when synthetic GPT-2 EEG signals are introduced during training. Similarly, the Random Forest classifying EMG data increases from 93.62% (0.8) to 93.9% (0.59) when training data is augmented by synthetic EMG signals. Additionally, as predicted, augmentation with synthetic biological signals also increases the classification accuracy of data from new subjects that were not observed during training. A Robotiq 2F-85 Gripper was finally used for real-time gesture-based control, with synthetic EMG data augmentation remarkably improving gesture recognition accuracy, from 68.29% to 89.5%.

I. INTRODUCTION

When presenting their Generative Pretrained Transformer (GPT) model, researchers at OpenAI hypothesised that *language models are unsupervised multitask learners* [1]. At the current state-of-the-art this claim has been consistently argued through applications such as fake news identification [2], patent claims [3], and stock market analysis [4] to name just a few in a rapidly growing area of research. In this work, we follow those before us in exploring the capabilities of these

models in a brand new field of application: the generation of bio-synthetic signals (in our case Electroencephalographic (EEG) and Electromyographic (EMG) activity). In detail, we aimed at exploring whether or not GPT-2's self-attention based architecture was capable of creating synthetic signals, and if those signals could improve the performance of classification models used on real datasets. Enabling better results for the deduction of a physical action or mental thought allows for a higher degree of certainty when it comes to an unseen subject. That is, for example in electromyographically controlled robotic prosthetic limbs, a more improved experience for the user of such a robotic device. Our scientific contributions and results suggest that:

1) It is possible to generate synthetic biological signals by tuning a language transformation model.
2) Classifiers trained on either real or synthetic data can classify one another with relatively high accuracy.
3) Synthetic data improves the classification of the real data both in terms of model benchmarking and classification of unseen samples.

II. RELATED WORK AND BACKGROUND

In this section, we describe how previous work has demonstrated the benefits of augmenting biological signal datasets to improve classification results, since it has been noted that augmentation is a useful technique to overcome data scarcity in such domains [5]. A common approach is to generate synthetic signals by re-arranging components of real data. Lotte [6] proposed a method of "*Artificial Trial Generation Based on Analogy*" where three data examples x_1, x_2, x_3 provide examples and an artificial $x_{synthetic}$ is formed which is to x_3 what x_2 is to x_1. A transformation is applied to x_1 to make it more similar to x_2, the same transformation is then applied to x_3 which generates $x_{synthetic}$[1]. This approach was shown to improve performance of a Linear Discriminant Analysis classifier on three different datasets. Dai et al. [7] performed similar rearrangements of waveform components in both the time and frequency domains to add three times the amount of initially collected EEG data, finding that this approach could improve the classification accuracy of a Hybrid Scale Convolutional Neural Network. This work showed that data augmentation allowed the model to improve the classification of data for individual subjects that were specifically challenging in terms of the model's classification ability. Dinarès-Ferran [8] decomposed EEG signals into Intrinsic Mode Functions and constructed synthetic data frames by arranging these IMFs into new combinations, demonstrating improvements of classification performance of motor imagery based BCIs while includ-

[1]Equations for Lotte's EEG generation technique can be found in [6]

ing these new signals. Other researchers have proposed data augmentation techniques commonly used in other domains such as image classification techniques with positive results. As an example Shovon et al. [9] applied conventional image augmentation techniques e.g. rotation, zoom, and brightness to spectral images formed from EEG analysis to increase the size of a public EEG dataset. This ultimately led to an improvement over the state-of-the-art. Current research shows great impact can be derived from relatively simple techniques. For example, Freer [10] observed that introducing noise into gathered data to form additional data points improved the learning ability of several models which otherwise performed relatively poorly. Tsinganos et al. [11] studied the approaches of magnitude warping, wavelet decomposition, and synthetic surface EMG models (generative approaches) for hand gesture recognition, finding classification performance increases of up to +16% when augmented data was introduced during training. More recently, data augmentation studies have begun to focus on the field of deep learning, more specifically on the ability of generative models to create artificial data which is then introduced during the classification model training process. In 2018, Luo et al. [12] observed that useful EEG signal data could be generated by Conditional Wasserstein Generative Adversarial Networks (GANs) which was then introduced to the training set in a classical train-test learning framework. The authors found classification performance was improved when such techniques were introduced. Likewise, Zhang and Liu [13] applied similar Deep Convolutional GANs (DC-GAN) to EEG signals given that training examples are often scarce in related works. As with the previous work, the authors found success when augmenting training data with DC-GAN generated data. Zanini and Colombini [14] provided a state-of-the-art solution in the field of EMG studies when using a DC-GAN to successfully perform *style transfer* of Parkinson's Disease to bio-electrical signals, noting the scarcity of Parkinson's Disease EMG data available to researchers as an open issue in the field [14]. Many studies observed follow a relatively simple train/test approach to benchmarking models.

A limitation of many techniques is that they are not temporal in their generative natures. Each block of signal output has no influence on the next, and, as such, a continuous synthetic signal of unlimited length cannot therefore be generated. Our approach allows for infinite generation of temporal wave data given the nature of GPT-2; a continuous synthetic raw signal is generated by presenting some of the previous outputs as input for the next generation. We then benchmark the models through k-fold cross validation, where each fold has synthetic data introduced as additional training data. Moreover, for the first time in the field, we show the effectiveness of attention-based models at the signal level rather than generative based models at the feature-level for both training and unseen data. We then finally show that real-time gesture classification towards direct control of a robotic arm is improved following our data augmentation framework.

A. GPT-2 and Self-Attention Transformers

Self-Attention Transformers are based on calculating scaled dot-product attention units, and generate new data by learning to paying attention to previous data generated [15]. Scaled dot-product attention is calculated for each unit within the input vector, e.g. words in a sentence, or, in this case, signals in a stream. The attention units are input with a sequence and output embeddings of relevant tokens. Query (W_q), key (W_k), and value (W_v) weights are calculated as:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (1)$$

where the query is an entity within the sequence, keys are vector representations of the input, and the values are derived by querying against keys. The term self-attention comes from the fact that Q, K and V are received from the same source, and generation is an unsupervised. GPT-2 architecture follows the concept of Multi-headed Attention:

$$MultiHead(Q,K,V) = Concat(head_1, ..., head_h)W^O$$
$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V). \quad (2)$$

That is, a deep structure of h_i attention heads in order to inter-connect multiple attention units. Fundamentally, the GPT and GPT-2 algorithms do not differ. The main advantages of GPT-2 are based on it being many times more complex than the GPT with 1.5 billion parameters and being trained on a large dataset of 8 million websites.

III. METHOD

A. Data Collection, Pre-processing and Feature Extraction

The EMG dataset used in this study was initially acquired by Dolopikos et al. in [16]. EMG data corresponding to the opening and closing movements of the right hand were collected from fifteen able-bodied participants (9 male, 6 female, mean age 26) using a Thalmic Labs Myo armband. The participants performed the gestures after a cue from an instructor. The recorded data corresponding to the time before the onset of physical activity (muscular background tone) was extracted and compiled into a third "neutral" class. To assess contraction and relaxation of muscles, information can be extracted by the simple analysis of an EMG signal's smoothed rectified envelope [17]. The data was indeed first rectified and then low-pass filtered using a peak detection algorithm [18], interpolating between local maxima with a separation of at least 20 samples (equivalent to 0.1 seconds at the Myo's natural sample rate of 200Hz). The EEG dataset used was initially acquired for a previous study [19]. A total of 5 participants were presented with stimuli while wearing the InteraXon Muse headband to collect EEG data from the TP9, AF7, AF8, and TP10 electrodes. EEG data corresponding to three mental states was collected from each participant: a neutral class with no stimulus present, relaxation enabled by classical music, and concentration induced by a video of the "shell game" (wherein they had to follow a ball placed underneath one of three shuffled upturned cups).

Whilst the data was provided to GPT-2 in its raw format, an ensemble of features was extracted from each dataset to enable classification. The feature set has previously proven effective, providing sufficient information to discriminate both between focused, relaxed, and neutral brains [19], and closed,
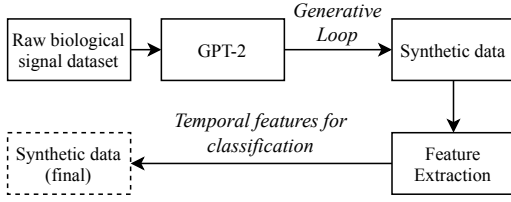
Fig. 1: Initial training of the GPT-2 model and then generating a dataset of synthetic biological signals.

open, and neutral hands [16]. Features are extracted from a sliding window of 1 second in length, at an overlap of 0.5 seconds. These windows are further sub-divided into halves and quarters, enabling extraction of the following ensemble of statistical features[2]:

- 1-second window
  - Mean $\bar{y}_k = \frac{1}{N} \sum_{i=1}^{N} y_{ki}$, and Standard deviation $s_k = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (y_{ki} - \bar{y}_k)^2}$ of the waveform
  - Skewness $g_{1,k} = \frac{\sum_{i=1}^{N} (y_{ki} - \bar{y}_k)^3}{N s_k^3}$, and Kurtosis $g_{2,k} = \frac{\sum_{i=1}^{N} (y_{ki} - \bar{y}_k)^4}{N s_k^4} - 3$ of each waveform
  - Maximum and minimum values over the given period
  - Sample variances of each wave, and sample co-variances of all unique pairs of waves $s_{k\ell} = \frac{1}{N-1} \sum_{i=1}^{N} (y_{ki} - \bar{y}_k)(y_{\ell i} - \bar{y}_\ell); \quad \forall \ k, \ell \in [1, K]$
  - Eigenvalues of the covariance matrix $det\left(\mathbf{S} - \boldsymbol{\lambda}\mathbf{I}_K\right) = 0$
  - Upper triangular elements of the matrix logarithm of the covariance matrix $e^{\mathbf{B}} = \mathbf{I}_K + \sum_{n=1}^{\infty} \frac{\mathbf{S}^n}{n!}$
  - The magnitude of each signal's frequency components, obtained via Fast Fourier Transform (FFT)
- 0.5-second windows
  - The change between the first and second sliding window in the sample mean and standard deviation and also the maximum and minimum values
- 0.25-second quarter windows produced due to offset
  - The mean of each signal in the 0.25-second window
  - All paired differences of means between the windows
  - Maximum and minimum values and their paired differences

B. Generating and Learning from GPT-2 Generated Signals

GPT-2 models are initially trained on each class of data for 1,000 steps each. Then, for n classes, n GPT-2s are tasked with generating synthetic data and the class label is finally manually added to the generated data. This process can be observed in Figure 1 where the generative loop is prefixed by the latter half of the previously generated data[3]. The synthetic equivalent of 60 seconds of data per class are generated (30,000 rows per class of raw signal data). To benchmark

[2]Feature extraction code available at
https://github.com/jordan-bird/eeg-feature-generation
[3]Example code can be found at: https://github.com/jordan-bird/Generational-Loop-GPT2

machine learning models, a K-fold cross validated learning process is followed and compared to the process observed in Figure 2 where training data is augmented by the synthetically-derived data at each fold of learning. The testing set does not contain any of the artificial signal data. This process is performed for both the EEG and EMG experiments for six different models: Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbours (KNN, $K = 10$), Linear Discriminant Analysis (LDA), Logistic Regression (LR), and Gaussian Naïve Bayes (GNB). These statistical models are selected due to their differing nature, to explore the hypothesis with a mixed range of approaches. As was explored in [20], it was found that unseen signal classification can be improved through calibration via inductive and supervised transductive transfer learning. That is, tuning a model by providing a small amount of calibration data to the training set.

IV. OBSERVATIONS AND RESULTS

In comparison, it was noted that all synthetic data was unique compared to the real data. A sample of real and synthetic EEG data can be observed in Figure 3. Interestingly, natural behaviours, such as the presence of characteristic oscillations, can be observed within data, showing that complex natural patterns have been generalised by the GPT-2 model. It is noted that in the real data, some spikes are observed in the signals from all electrodes but those are likely due to involuntary (and unwanted) eye blinks. Worth nothing is that the GPT-2 does not replicate similar patterns, most likely as a filtering side-effect of data generalisation, since such occurrences are random and unrelated to the underlying EEG data. The Power Spectral Densities of the GPT-2 generated data were computed with Welch's method [21] and compared with those computed from real human data as can be seen in Figure 5. In observing the frequency domain plots of the genuine data, there is a clear 50Hz component in all classes likely due to power-line interference. Interestingly, there has been a clear attempt by GPT-2 to mimic this feature, albeit with a much shallower roll-off. Figure 4 shows the same process for EMG data, where the GPT-2 generated waves are seemingly less natural than their human counterparts; although natural wave patterns do emerge, they are more erratic and prone to spiking unlike the signals recorded from a human forearm. The Power Spectral Densities presented in Figure 6 indicate that across all classes the synthetic data has significantly more power in its high frequency components than the real data. Despite the real EMG dataset having been low-pass filtered before being used to train GPT-2 this phenomenon is more notable in the EMG domain, due likely in part to the aforementioned erratic nature of the synthetic EMG signals.

A. Classification of real-to-synthetic data and vice-versa

Table I shows the effects of training models on the real and synthetic EEG data and then attempting to classify the other data. Interestingly, the Support Vector Machine when trained on real data can classify the synthetic data with 90.84% accuracy. Likewise, the Gaussian Naïve Bayes approach when trained on the synthetic data can then classify the real data
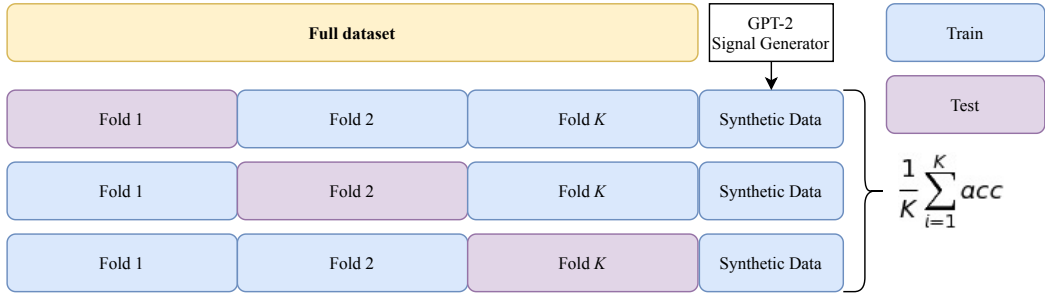
Fig. 2: The standard K-Fold cross validation process with the GPT-2 generated synthetic data being introduced as additional training data for each fold.
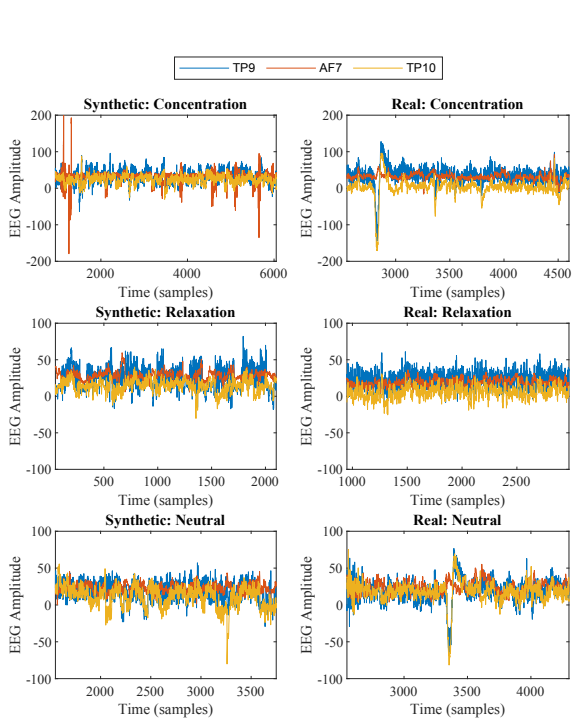

Fig. 3: Comparison of GPT-2 generated (Left) and genuine recorded (Right) EEG data across "Concentrating", "Relaxed", and "Neutral" mental state classes. AF8 electrode readings are omitted for readability purposes.
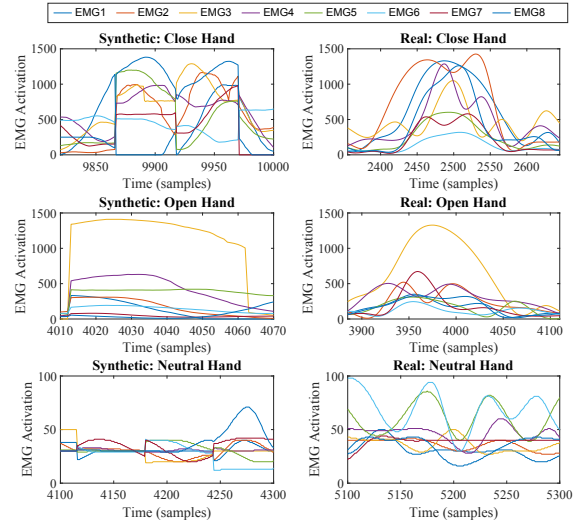

Fig. 4: Comparison of GPT-2 generated (Left) and genuine recorded (Right) EMG data across "Closed", "Open", and "Neutral" hand classes.

TABLE II: Classification results when training on real or synthetic EMG data and attempting to predict the class labels of the other (sorted for real to synthetic).

Classifier	Training and Prediction Data	
	Real to Synthetic	Synthetic to Real
Gaussian Naïve Bayes	**62.36**	64.39
10 Nearest Neighbours	62.07	**78.24**
Random Forest	61.78	71.23
Linear Discriminant Analysis	50.00	60.69
Logistic Regression	37.36	71.71
Support Vector Machine	35.63	71.27

with 74.71% accuracy. Table II similarly shows the ability to classify real data by learning from synthetic data and vice versa for EMG. The NB model when trained on only real data can classify the synthetic data with 62.36% accuracy, whereas the KNN model can classify the real dataset with 78.24% accuracy when trained on only synthetic.

B. EEG Classification

The results for EEG classification can be seen in Table III. The best result overall for the dataset was the k-fold training

TABLE I: Classification results when training on real or synthetic EEG data and attempting to predict the class labels of the other (sorted for real to synthetic).

Classifier	Training and Prediction Data	
	Real to Synthetic	Synthetic to Real
Support Vector Machine	**90.84**	66.88
Random Forest	88.14	70.71
10 Nearest Neighbours	85.18	72.13
Linear Discriminant Analysis	77.90	68.90
Logistic Regression	70.22	64.91
Gaussian Naïve Bayes	67.52	**74.71**

Fig. 5: Comparison of Power Spectral Densities of GPT-2 generated (Left) and genuine recorded (Right) EEG data. For readability, only the PSD computed from electrode TP9 is shown.
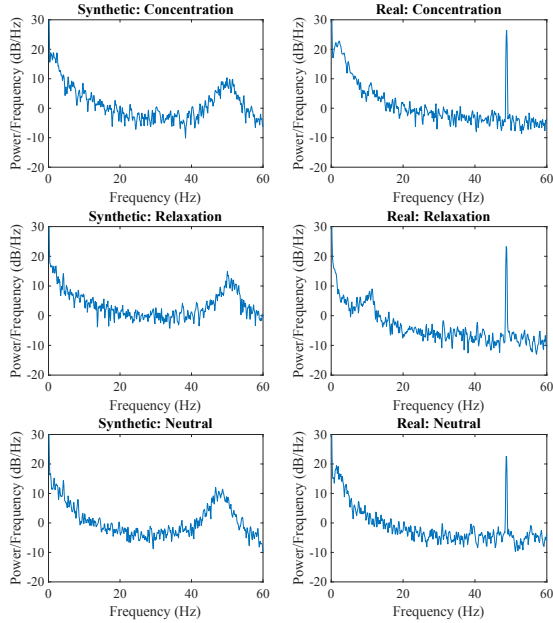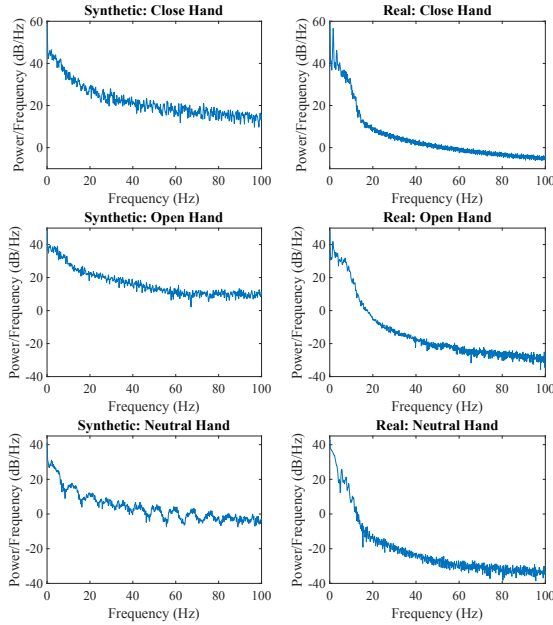

Fig. 6: Comparison of Power Spectral Densities of GPT-2 generated (Left) and genuine recorded (Right) EMG data. For readability, only the PSD computed from electrode EMG1 is shown.

TABLE III: Comparison of the 10-fold classification of EEG data and 10-fold classification of EEG data alongside synthetic data as additional training data.

Classifier	Without GPT-2	With GPT-2 Data
Random Forest	95.81 (1.46)	**96.69 (1.12)**
Logistic Regression	93.71 (1.05)	93.30 (1.37)
Support Vector Machine	93.67 (1.35)	93.71 (1.33)
Linear Discriminant Analysis	91.93 (1.24)	94.03 (1.29)
10 Nearest Neighbours	89.83 (1.75)	90.68 (2.07)
Gaussian Naïve Bayes	70.27 (2.53)	72.41 (2.33)

TABLE IV: EEG classification abilities of the models on completely unseen data with regards to both with and without synthetic GPT-2 data as well as prior calibration.

Classifier	Uncalibrated		Calibrated	
	Vanilla	*Synth.*	*Vanilla*	*Synth.*
Random Forest	38.84	42.90	59.75	59.98
Logistic Regression	46.35	47.01	46.92	48.10
Support Vector Machine	47.11	47.00	53.45	52.80
Linear Discriminant Analysis	56.07	57.48	63.85	**66.02**
10 Nearest Neighbours	48.29	48.78	59.64	60.60
Gaussian Naïve Bayes	48.25	48.97	49.62	50.37

process with additional training data in the form of GPT-2 generated synthetic brainwaves, using a Random Forest. This achieved a mean accuracy of 96.69% at a deviance of 1.12%. Table IV shows the classification abilities of the models when given completely unseen data from three new subjects. The results show the difficulty of the classification problem faced, with many scoring relatively low for the three-class problem. The best result was found to the the Linear Discriminant Analysis model when trained with both calibration and synthetic GPT-2 data alongside the dataset, which then scored 66.02% classification accuracy on the unseen data.

C. EMG Classification

Table V shows the results for EMG classification. The best model was the Random Forest which scored 93.9% (deviance 0.59) during the k-fold benchmarking process in which GPT-2 synthetic data was introduced as additional training data. Table VI shows the abilities of the models when predicting the class label of completely unseen EMG data. Interestingly, the Gaussian Naïve Bayes model outperformed all others consistently. The best Gaussian Naïve Bayes model at predicting completely unseen data was when it was also trained with

TABLE V: Comparison of the 10-fold classification of EMG data and 10-fold classification of EMG data alongside synthetic data as additional training data.

Classifier	Without GPT-2	With GPT-2 Data
Random Forest	93.62 (0.8)	**93.90 (0.59)**
Logistic Regression	93.75 (1.04)	93.86 (1.05)
Support Vector Machine	93.42 (0.89)	93.46 (0.94)
Linear Discriminant Analysis	91.95 (1)	92.59 (1.05)
10 Nearest Neighbours	91.23 (0.89)	91.11 (0.88)
Gaussian Naïve Bayes	77.73 (1.39)	74.46 (1.38)

TABLE VI: EMG classification abilities of the models on completely unseen data with regards to both with and without synthetic GPT-2 data as well as prior calibration.

Classifier	Uncalibrated		Calibrated	
	Vanilla	*Synth.*	*Vanilla*	*Synth.*
Random Forest	67.33	69.31	74.26	75.25
Logistic Regression	60.40	87.13	60.40	87.13
Support Vector Machine	39.60	62.38	44.55	46.53
Linear Discriminant Analysis	65.35	67.33	86.14	79.21
10 Nearest Neighbours	75.25	75.25	78.22	78.22
Gaussian Naïve Bayes	95.05	94.06	96.04	**97.03**
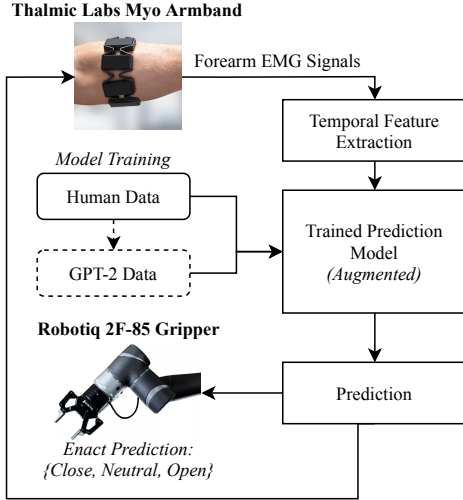
Fig. 7: Real-time predictions of EMG signals enacted by the Robotiq 2F-85 Gripper.

Fig. 8: Real-time execution of gestures for three minutes predicted with the augmented EMG model (89.5%) and non-augmented EMG model (68.29%).

Fig. 9: Confusion Matrix for real-time EMG classification
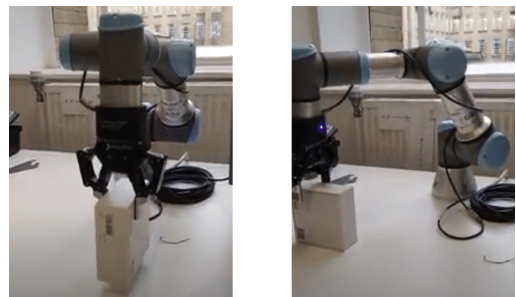
Fig. 10: A Universal Robotics UR3 Manipulator and Robotiq 2F-85 gripper picking up and then releasing an object.

calibration and GPT-2 synthetic data alongside the dataset at an accuracy of 97.03%.

D. Real-time EMG Prediction for the Control of a Robotic Manipulator

The overall process followed for robotic enaction of predicted hand gestures by a Robotiq 2F-85 Gripper can be seen in Figure 7. The results in Figure 8 show the process of a user performing hand gestures for three minutes (124 data objects). The best-performing EMG prediction model was applied (Gaussian Naïve Bayes + GPT-2), which predicted real-time data with 89.5% accuracy. All of the erroneous predictions occurred during state transitions, which was expected given that models were trained on concrete gestures and had not been exposed to transitional behaviours of the arm muscles when shifting between gestures. The best predictive model on the dataset without GPT-2 augmentation scored 68.29% accuracy. The 95% Wilson confidence interval for the augmented model's accuracy was [82.89, 93.77], and for the non-augmentation model was [59.62,75.86]. No calibration was performed, that is, the models were never exposed to data from this user. Thus, GPT-2 biosignal data augmentation leads to a model which can classify data from unseen subjects with a higher rate of success. Figure 9 shows the confusion
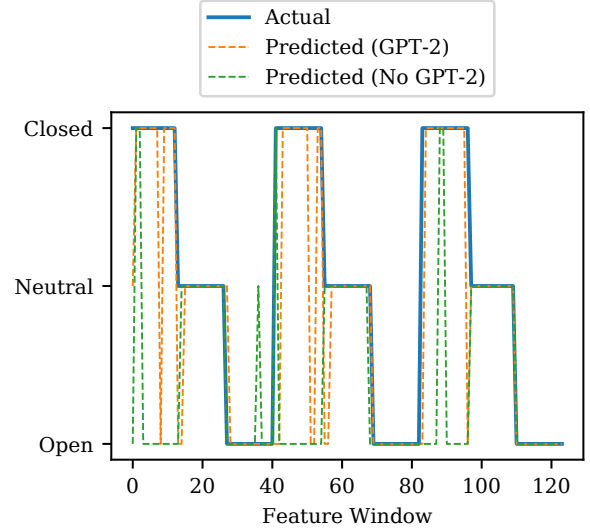
matrix for this experiment. An application of the approach is shown in Figure 10, where a pick-and-place routine and the EMG classifier control a UR3 Manipulator's Robotiq 2F-85 gripper [22]. The device mimics the user and allows for teleoperation in order to pick up (grip) and place (release) an object. If the operator keeps their hand in a neutral position, then no movements are commanded to the artificial arm.

V. CONCLUSIONS AND FUTURE WORK

To conclude, this study has presented multiple experiments with real and synthetic biological signals in order to ascertain whether classification algorithms can be improved by considering data generated by the GPT-2 model. Although the data are different, i.e., real and synthetic data were unique, a model trained on one of the two sets of signals can strongly classify the other and thus the GPT-2 model is able to generate relatively realistic data which holds useful information that can be learnt from for application to real signals. For EEG, an SVM trained on synthetic data could classify real data at 74.71% accuracy and a KNN algorithm could do the same for real EMG classification at 78.24% accuracy, training on only synthetic data. We then showed that several learning algorithms were improved for both EMG and EEG classification when the training data was augmented by GPT-2. The main argument of this work is that synthetic biosignals generated by an attention-based transformer hold useful information towards improving several learning algorithms for classification of real biological signal data. In future, larger datasets could be used and thus deep learning would be a realistic possibility for classification following the same process. Given that this work showed promise in terms of the model architecture itself, similar models could also be benchmarked in terms of their ability to create augmented training datasets e.g. BART, CTRL, Transformer-XL and XLNet. Another unoptimised level of detail is the amount of synthetic data that is added to the training set for augmentation, future work could explore the level of data needed for apt improvements to the models.

Our suggested model for EMG, the GNB approach trained with human-sourced GPT-2 generated synthetic signals, was powerful in terms of predictive ability and required relatively little computational resources given its simplistic nature. Additionally, the approach did not require further calibration, as many state-of-the-art approaches do (including the Myo software itself), instead correctly predicting the behaviours of a new subject from the point of wearing the device. Given these attributes, the model is apt for usage on-board within wearable EMG devices for real-time prediction of gesture.

REFERENCES

[1] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI Blog*, vol. 1, no. 8, p. 9, 2019.

[2] J. C. B. Cruz, J. A. Tan, and C. Cheng, "Localization of fake news detection via multitask transfer learning," *arXiv preprint arXiv:1910.09295*, 2019.

[3] J.-S. Lee and J. Hsiang, "Patent claim generation by fine-tuning openai gpt-2," *arXiv preprint arXiv:1907.02052*, 2019.

[4] Y. Nishi, A. Suge, and H. Takahashi, "Text analysis on the stock market in the automotive industry through fake news generated by gpt-2," *Proceedings of the Artificial Intelligence of and for Business*, 2019.

[5] E. Lashgari, D. Liang, and U. Maoz, "Data augmentation for deep-learning-based electroencephalography," *Journal of Neuroscience Methods*, p. 108885, 2020.

[6] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 871–890, 2015.

[7] G. Dai, J. Zhou, J. Huang, and N. Wang, "HS-CNN: a CNN with hybrid convolution scale for EEG motor imagery classification," *Journal of Neural Engineering*, vol. 17, jan 2020.

[8] J. Dinarès-Ferran, R. Ortner, C. Guger, and J. 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, p. 308, 2018.

[9] T. H. Shovon, Z. A. Nazi, S. Dash, and M. F. Hossain, "Classification of motor imagery eeg signals with multi-input convolutional neural network by augmenting stft," in *2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*, pp. 398–403, 2019.

[10] D. Freer and G.-Z. Yang, "Data augmentation for self-paced motor imagery classification with c-LSTM," *Journal of Neural Engineering*, vol. 17, jan 2020.

[11] P. Tsinganos, B. Cornelis, J. Cornelis, B. Jansen, and A. Skodras, "Data augmentation of surface electromyography for hand gesture recognition," *Sensors*, vol. 20, no. 17, p. 4892, 2020.

[12] Y. Luo and B.-L. Lu, "Eeg data augmentation for emotion recognition using a conditional wasserstein gan," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2535–2538, IEEE, 2018.

[13] Q. Zhang and Y. Liu, "Improving brain computer interface performance by data augmentation with conditional deep convolutional generative adversarial networks," *arXiv preprint arXiv:1806.07108*, 2018.

[14] R. Anicet Zanini and E. Luna Colombini, "Parkinson's disease emg data augmentation and simulation with dcgans and style transfer," *Sensors*, vol. 20, no. 9, p. 2605, 2020.

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, pp. 5998–6008, 2017.

[16] C. Dolopikos, M. Pritchard, J. J. Bird, and D. R. Faria, "Electromyography signal-based gesture recognition for human-machine interaction in real-time through model calibration," in *SAI Future of Information and Communication Conference (FICC) 2021*, 2021.

[17] C. J. D. Luca, "The use of surface electromyography in biomechanics," *Journal of Applied Biomechanics*, vol. 13, no. 2, pp. 135–163, 1997.

[18] The MathWorks, Inc., *Signal Processing Toolbox^{TM} Reference*, pp. 489–503. 2020.

[19] J. J. Bird, L. J. Manso, E. P. Ribeiro, A. Ekart, and D. R. Faria, "A study on mental state classification using eeg-based brain-machine interface," in *2018 International Conference on Intelligent Systems (IS)*, pp. 795–800, IEEE, 2018.

[20] J. Kobylarz, J. J. Bird, D. R. Faria, E. P. Ribeiro, and A. Ekárt, "Thumbs up, thumbs down: non-verbal human-robot interaction through real-time emg classification via inductive and supervised transductive transfer learning," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–11, 2020.

[21] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[22] Universal Robotics, "Ur3 technical specifications." Technical Specifications, march 2015.