Strategies for Efficient ENG Signal Classification using Data Augmentation and Data Balancing Techniques

Antonio Coviello[1*], Rafael Sobral Augusto[1*], Davide Gottardello[1] , Matteo Esposito[1], Matteo Manetti[1], Umberto Spagnolini[1], Maurizio Magarini[1]

[1] Department of Electronics, Information and Bioengineering, Politecnico di Milano, Piazza L. Da Vinci 32, 20133 Milan, Italy
[*]Both authors contributed equally to this research.
antonio.coviello@polimi.it

***Abstract*— Peripheral nerve injuries interrupt essential brain-body communications, leading to significant functional impairments. This study evaluates five Convolutional Neural Network (CNN)-based classification methods for Sciatic Nerve Electroneurographic (ENG) signals in rats, aiming to restore the lost connection. The examined networks include classic CNN, Convolutional Tiny Transformer (CTT), InceptionTime (IT), IT with Derivative (IT-D), and ENG Network (ENGNet). Light preprocessing ensures real-time application, essential for human sensory systems, with a maximum delay of 300ms. Different approaches of data augmentation and data balancing strategies are used to address dataset imbalances. These approaches prove to be optimal for rebalancing, particularly the overlapping techniques, which significantly enhance the classification of the previously underestimated classes.**

Keywords— ENG signal, Data Augmentation, Data Balancing, ENGNet.

I. Introduction

Peripheral nerve lesions represent a clinical problem whose incidence on the population is constantly increasing. Aging and lifestyle have contributed significantly to increase their frequency. When an individual suffers a peripheral nerve injury, the body's functionality is impaired, interrupting communications between the brain and various parts of the body. These injuries generate troubles, such as permanent disability or impediment to daily activities, preventing the affected person from living a normal lifestyle. For this reason, the clinical picture mentioned requires finding a practical solution that is able to cure this type of impairment [1,2,3].

In the field of neuroprosthetics, a noteworthy progress concerns the creation of implantable devices. The devices are designed to bypass the damaged nerve sites, creating a neural bypass to recover lost function. One promising approach involves using a closed-loop interface that measures Electroneurography (ENG) signals in the distal (or proximal) nerve portion of the lesion, with the goal of restoring patient sensation (or movement) bypassing the injured site. The implementation of this alternative approach to traditional medicine would allow information to be transmitted despite the presence of a lesion. Treatment and restoration of lost activity would be guaranteed to all types of patients suffering from this pathology [1, 2, 4].

The signal would be sampled from the region of interest via a cuff electrode, using the implanted device. For these measurements to be effective, a classification procedure must be performed to identify which stimuli were received by the subject and transmit the correct information upstream into the nervous system once the sensory input has been decoded. The extracted information would be used to carry out neuromodulation to restore the connection [1, 4].

Despite the high effort put into research in recent years, there are still many gaps associated with ENG signal classification [1]. Within this work, different types of classifiers based on Convolutional Neural Networks (CNN) will be analyzed to classify the ENG signals acquired from the sciatic nerve of rats. The aim is to understand how to carry out signal classification efficiently by taking into consideration different types of problems exploiting different classifiers. The issues here addressed concern:

- Number of epochs needed to train the classifier quickly and efficiently;
- Limited amount of data available to train the classifier;
- Class imbalance, as some signals may appear more frequently than others;
- Time constraint, so that the classification have to be done within a confidence interval below the human perception time;
- Limited memory capacity of the implanted device.

The networks treated for the analysis are of different types, but they are all based on the use of kernels that, by scrolling within the signal, allow the information contained in the ENG signal to be extracted and classified. The first network used is a classic CNN. Using 2 sets of 1D kernels, across 6 repeated convolutional blocks, it is able to extract temporal and spatial information from the signal. The second is the Convolutional Tiny Transformer (CTT) based on a temporal and spatial attention mechanism to process the input by generating output based on the encoded information. The third is InceptionTime (IT), based on the use of inception modules for time series classification. The network behaves like a bank of filters that selects information specific to the size of the window used. The fourth is the IT-Derivative (IT-D), which has the same structure as IT, but instead of containing only 3 low-pass filters, it includes a fourth derivative high-pass filter. This allows us to consider a broader frequency spectrum. Finally, the ENG Network (ENGNet) is analyzed that, using a combination of convolutional layers and clustering, is able to maximize the extraction of temporal and spatial patterns, thus proving to be the most efficient among all the considered ones. More information about CNN, IT, IT-D, and ENGNet are given in [4]. The CTT was inspired from [5] and is explained in Sec. III.

II. DATASET AND PREPROCESSING

The analysis covered in this paper was conducted using a database available in [6]. A second dataset reported in [7] has been used to verify the trend of the results. Both datasets exhibit similar trends. In this paper, we present results only from the first dataset [6]. To capture ENG signals, mechanical stimulations were administered to the limbs of three Sprague Dawley rats. An Arduino-based experimental setup was used to standardize stimulation levels, as detailed in [6]. Three categories of somatosensory stimuli were implemented on the right hind paw of rats: nociception, proprioception (including dorsiflexion and plantarflexion stimuli), and tactile sensation.

Nociception was elicited by manually pinching each rat's hind paw with forceps. The procedure was repeated 50 times on both the nail and the heel. Each stimulus has an average duration of 1 second followed by 3 seconds of inactivity. To examine proprioceptive sensations, the rats' nails were attached to a bar connected to a servo motor, minimizing tactile stimulation. The limb was held at a constant angle of approximately 70º, and the bar was moved to three distinct angles (10º, 20º, and 30º) in both directions. Each angle was repeated 50 times per subject. Each stimulus has an average duration of 3 seconds followed by 3 seconds of no activity. Finally, tactile stimulation was achieved using two Von Frey (VF) fibers with a force of 100 and 300 grams, manipulated by a linear motor. This procedure was also repeated 50 times. Each stimulus has an average duration of 3 seconds followed by 6 seconds of no activity. For the case of ENGNet, a fifth class was added using all the rest period from all the signals. The touch signal, instead, it is not considered in the Rest class due to his mix behavior during rest period with proprioception.

ENG signals were acquired using multi-contact sleeve electrode placed in the distal part of the rat sciatic nerve, with a sampling rate of 30 kHz. The electrode was composed of 16 channels, 4 rings with 4 contacts each, allowing to obtain spatial information in the same location of the nerve and the progression of the signal over time. Classification was performed on four distinct types of activity: dorsiflexion, plantarflexion, touch, and pain. To ensure precise synchronization between stimulus and measured data, sensors were applied that allowed the exact start of stimulation to be measured. The first and last 0.25 seconds of each stimulus were removed to eliminate any transitional effects between the non-stimulation period and the stimulus period. Only static intervals were used for classification.

Very light preprocessing has been carried out to obtain results in real-time applications [4]. This is because the maximum non-perceptible delay for human sensory systems is around 300 ms [8]. Consequently, motor or sensory restoration, starting from sampling up to neuromodulation, must fall within this time frame. This would ensure a natural sensation to the subject without perceiving a deficit. The ENG signals were pre-processed first with filtering [4]. An eighth-order Butterworth bandpass filter in the range 0.8-2.5 kHz was subjected to the signals to remove the low-frequency contribution of EMG and other types of high-frequency noise [9]. A downsampling was subsequently carried out. The signals were reduced to 5 kHz to reduce the computational load and preserve the main characteristics for signal classification. As reported in [10] this step has no significant effect on the classification of the signal. Due to the high voltage peaks present in the raw signals, generated by external elements, an empirical thresholding operation was performed.

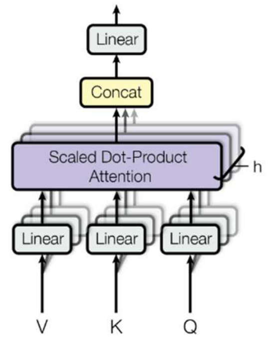


Figure 1. Schematic representation of multihead attention mechanism [12].

The physiological threshold was set to 30 μV to eliminate measurements whose absolute value exceeded this value. This threshold led to the deletion of less than 0.1% of the total data. Finally, the signal was divided into non-overlapping windows of different sizes to proceed with the classification. The window sizes were chosen to allow real-time operations of the investigated classifiers: 500 ms, 200 ms, 100 ms, and 50 ms [11]. Each model is trained individually for each input window, using a matrix as input, with the window size as the length and the number of electrodes as the rows.

III. DATA CLASSIFICATION

In this work, five types of networks are used to implement the ENG signal classification. For four of these, i.e. CNN, IT, IT-D, and ENGNet more information is reported in [4], while CTT is here introduced by takin inspiration from [5]. CTT is differs from other models as it relies mainly on the attention mechanism [12] and not only on convolution. As depicted in Fig. 1, the multihead attention mechanism of the CTT is made by three characteristic blocks: Spatial Transforming (ST), Positional Embedding (PE), and Temporal Transforming (TT). ST utilizes a method inspired by scaled dot product attention to weight signal's channels, optimizing their use by prioritizing significant features and using residual connections for gradient stability. PE uses convolution to encode positional information, compressing data along channels, and segment them in shorter slices. Finally, TT uses multi-head attention to capture global temporal dependencies in the slices, enhancing understanding of temporal dynamics. At the end, a fully connected layer and a Softmax layer are added to translate the features extracted into predictions amongst the four classes in the dataset. Once the windowing of the signal was carried out, the windows obtained were used as input for the described neural networks. A separate network is created for each rat and for each considered window length. During the training for optimizing the hyperparameters of the classifiers, the risk of overfitting on the dataset used is always present. This risk arises because hyperparameters are adjusted until the model shows optimal performance with the data used, without considering any intrinsic variability in the data. To mitigate this problem, the dataset is divided into three subsets called training, validation and test sets. The model is trained on the training set, then the best model is selected based on the validation set and finally evaluated on the test set. However, this split into three sets reduces the amount of data available for learning the model. This can introduce variability due to the random selection of training and validation sets. Depending on the initial selection, different results will therefore always be obtained.

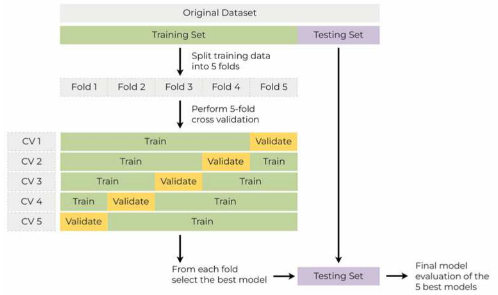


Figure 2. Scheme for correct training of classifiers.

Therefore, to address this problem, the Cross-Validation (CV) technique is commonly used. The data preparation and model evaluation process involve the following key steps, shown in Fig. 2. The complete dataset is initially divided into two distinct subsets: the learning and the testing set, which consists of 80% of the data for learning and the remaining 20% for the testing. A variation during the k-fold training, called StratifiedKFold [13], is used to return a statistical significance to the result: each subset contains approximately the same percentage of samples of each target class as the full set. The k-fold cross-validation procedure is performed exclusively on the model's learning set, which is divided into training and validation. In k-fold cross-validation, the training set is divided into k smaller subsets or "folds". For each fold, the model is trained on k−1 folds and validated on the last remaining fold. This process is repeated k times, each time with a different fold as the validation set. The best performing model, determined by its accuracy on the validation set during the k-fold cross-validation process, is selected. The "best" model is then used to make predictions on the test data, which were completely excluded during training and validation. This final step provides a new and reliable estimate of the model's performance on unseen data. K-fold cross-validation offers a solution that prevents overfitting on the test set by efficiently using available data. It is a powerful technique, especially useful in small sample scenarios, as it maximizes the use of data. In this work k=5 was used and the final performance was calculated with accuracy and F1-score, both considered as mean ± standard deviation over 5 folds. Furthermore, to prevent overfitting and improve the generalization capabilities of a model, a technique called Early Stopping (ES) was implemented. By monitoring the model's performance during training, when it does not improve for a certain number of iterations, training is stopped. Once the stopping criteria are reached, the training stops and the best epoch results are used as the final results. The ES model must ensure that at least 20 training epochs have been completed before stopping. The value was chosen empirically as the minimum necessary to obtain results above 90% accuracy for the input window considered. If no improvement is detected after the predefined number of epochs, training is considered complete and the final metrics are calculated and saved. In our study the ES was set after the eighth epoch and the interruption of training occurs only after 12 consecutive epochs without any improvement in validation accuracy.

IV. Data Augmentation and Data Balancing

Since the datasets have limited data and strong class imbalance, as reported in Fig. 3, several Data Augmentation (DA) and Data Balancing (DB), strategies were implemented to identify the optimal pipeline to consider. Both DA and DB approach can be applied to all ENG dataset that present this type of problems.

(a)

(b)

Figure 3. Data unbalancing, a) case 4 classes, b) case 5 classes.

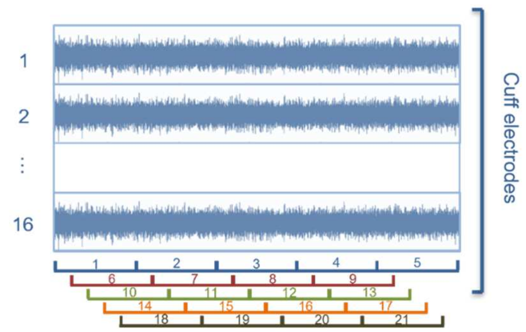


Figure 4. Data overlapping 80%.

The DA strategy adopted involves the superposition of the samples acquired from the original signals during a period of activity. This approach allows us to provide many more samples to the classifier, improving the overall accuracy and F1-score of the algorithm. It was decided empirically to overlap 80% of the samples to maximize the information extracted from each useful window, as illustrated in the Fig. 4. With these overlaps, three different classification strategies were used. The test set remained unchanged across all training set configurations and was consistently used in the same manner as in the non-overlapping case, as illustrated in Fig. 2. The subdivision of the dataset always occurred in the same way, eliminating the samples from the training data that overlapped with those selected in the test data for each of the CV training sets. This allowed us to ensure that the presence of overlap did not influence the accuracy of the classifier. The three strategies considered and tested were [14]:

1. Keep all overlapping samples in the training set: This strategy maximizes the total amount of information provided to the classifier by excluding from the training set only samples that had any overlap with the test set. However, data imbalance remained a factor during training.
2. Use the original samples from the class with the most data and add randomly selected overlapping samples only in all

other classes: this process discards many overlapping samples but allows us to have a balanced dataset with the same amount of samples for each class. However, since one of the classes has no overlap between its samples, this may impact the training process.

3. Add 10% of the size to the class with the most randomly selected data and balance all other classes: this process discards many overlapping samples but ensures that a balanced dataset is presented to the classifier. It also ensures that all the classes have at least 10% overlapping samples within the training dataset, so the lack of overlapping samples does not significantly impact the classifier's performance.

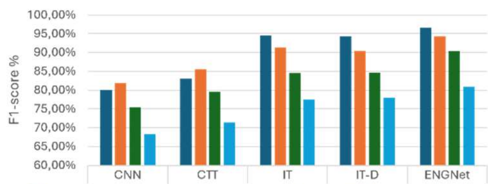
As far as the imbalance between the different classes is concerned, the DB was also tested using Random Undersampling, Random Oversampling, and Class Weights techniques. These strategies allow us to artificially change the total number of samples present in the classes by randomly selecting existing samples and duplicating or excluding them, respectively, within the training set. Once again, the test set was kept unchanged throughout all tests, while these mechanisms were only applied to the training set, both without overlapping but duplicating or reducing the data. In detail, the three strategies include [15, 16, 17, 18]:

1. For Random Undersampling, the minority class was kept unchanged while all other classes were reduced to a total of samples that was double the value of the smallest class or their total, if less than the first option. This was done to make the dataset more homogeneous compared to the class that had fewer samples, without losing further information.
2. For Random Oversampling, two methods were tested: the first, to oversample all classes up to the total present in the largest class, except in the case where the class has less than half of this value. In this case, only up to half of the total is oversampled. In the second, all classes are upsampled to the same value. Since there is a large difference between the smallest class and the others, increasing the total samples too much by copying them could actually have a negative effect on this class, leading to overfitting and endangering the network's ability to generalize its learning.
3. Finally, a Class Weights strategy was implemented [19]. It is characterized by the loss function used in the classification process, in order to perform a weighted average between the components of each class. This is done so that smallest classes, which would have less importance in the overall prediction, can have a similar weight in the final loss value when calculating the performance of the algorithm at each epoch. While it does not change the total samples of any class, it helps balance the prediction process by using larger weights for smaller classes and vice versa.

V. ANALITIC RESULTS

This section will report the numerical results obtained during the analysis described above. Figure 5 shows the results obtained for F1-score. This metric will be the one used as an evaluation criterion as it is the most stable in the presence of an unbalanced dataset, as in our case. It also allows us to highlight the presence of any false positives or negatives [4]. The five networks analyzed were compared and among these the ENGNet was found to be the best performing one. As can be seen, longer windows led to better results for all networks as there is a higher information contribution. By reducing the window the information content collapses. However for real-time implementation it is necessary to use small windows, which allow us to remain within the time confidence of 300ms [8]. Windows of 100 ms in the case of ENGNet have reported F1-score values above 90%, indicating the network as optimal for real-time applications. However, the two ITs also reported good values above 84%. By implementing DA techniques these could increase their performance making them equivalent candidates to ENGNet. Furthermore, both CTT and CNN were not suitable for a real-time application due to their poor results for low window values (79% and 76% F1-score respectively for 100ms window). In addition to the F1-score parameters, it is necessary to consider two other elements for real-time applications, the memory limitation of the implanted devices and the processing time to implement neuronal bypass. The fewer parameters there are, the easier the algorithm can be loaded onto the device, also reducing the power required. The number of parameters used per network is 577 956 for CNN, 50 948 for IT, 50 948 for IT-D, 5 796 for ENGNet and 5 484 for CTT. Considering that an implanted device has a typical memory of the order of 100 KB [22], and that a code must be loaded in order to make the device work, to guarantee redundancy in the memory the network must have a number of parameters below 10,000 otherwise it cannot be implemented.

The networks were found to be optimal in terms of number of parameters were CTT and ENGNet. Alternatively, in the case of a high number of parameters a secondary external device must be used to carry out a classification externally to the body. This allows us to eliminate the problems associated with the memory of the implanted part but introduces new ones associated with the transmission of data between the two devices [20]. The two IT networks could be implemented in this second solution. The IT-D network does not have additional parameters since with the derivative filter it performs a simple subtraction operation, slightly increasing its performance. As regards the time parameter, a qualitative study was carried out which could indicate which of these classifiers is the fastest. Depending on the microcontroller, the performance may change but the speed order will remain unchanged. The time required to classify 100 ms windows was 4.33ms for ENGNet, 2.71ms for CNN, 1.24ms for IT, 1.24ms for IT-D, and 0.07ms for CTT. The value introduced is exclusively associated with the classification time. To this must be added all the times necessary to carry out the neuronal bypass. For more information see [4]. Interestingly, despite the low F1-score values, the CTT network is the best in terms of number of parameters and classification execution time.



	CNN	CTT	IT	IT-D	ENGNet
500 ms	80,00±14,11	83,10±7,23	94,52±4,04	94,26±5,55	94,87±6,73
200 ms	81,82±10,11	85,50±4,23	91,39±6,51	90,49±7,78	94,30±6,30
100 ms	75,40±9,23	79,60±2,03	84,53±9,71	84,63±11,08	90,94±8,52
50 ms	68,19±8,82	71,40±1,27	77,52±12,56	77,89±12,63	80,39±9,11

Figure 5. Neural Network F1 score comparison.

A. DA and DB strategy results

To understand how the performance of the prediction process changes according to the length of the input, an analysis was conducted on the loss value during training, as illustrated in Fig. 6. From the figure, it can be clearly seen that longer input window require more epochs for the network to learn all the information present there, while the smaller input window reach a plateau much faster than the 40 epochs considered, also reaching overfitting and losing generalization capabilities if training is maintained for too long. This behavior was observed in all algorithms considered. The figure shows the behavior exclusive to CNN. This led to the development of the ES algorithm described above, consequently eliminating the problems associated with model overfitting.

Regarding the problems associated with the low amount of data and the unbalanced dataset, the DA and DB strategies were developed, compared, and reported in Fig. 7. The F1-score values obtained per 100ms window using the CNN are reported. The same trend was observed for all the networks. The results show that overlap 1, which consisted of adding all possible overlapping samples to the training set, gave the best results. In particular, by comparing the original results with those obtained using overlap 1, it is possible to observe a slight improvement of 5% for the three classes of plantarflexion, dorsiflexion, and touch, compared to an increase of almost 20% for the pain stimulus. This showed that an increase in data led to the improvement of classification consistently. In particular, the cause responsible for the improvement in performance was the 5-fold increase in data for the pain class. This indicates that in future experiments greater attention must be paid to signals that present difficulties during classification. However, as far as DB is concerned, none of the techniques significantly improved the overall prediction ability of the networks. However, the subsampling results were a surprise. By reducing the input size to almost half of the total samples, the results were still comparable to the original setup and other balancing strategies. This indicates that even if a smaller amount of data is used, with a balanced dataset, the same performance can be achieved without the use of overlap. This is important because it allows us to greatly reduce the time spent during the training phase of a classifier. At the same time, all strategies, both DA and DB, improved the prediction of Nociception, which is the smallest class, without compromising the prediction of the other classes. In cases where a classifier exhibits similar behavior, not classifying all classes evenly, overlap 1 proves to be the optimal solution. The results show that the classification of the other classes is not compromised, while the prediction of the weak class is enhanced. This behavior is constant for all the networks studied, in all animles in both dataset study.

Since the overlapping 1 technique was very satisfactory, the results associated with the DA technique are shown in the Fig. 8. It is interesting to observe how for 100 ms windows networks such as ENGNet and the two ITs obtain values higher than 95 % of F1-score. This indicates that, in particular for the two ITs, a larger dataset is needed to correctly train the classifiers. In addition, it is interesting to point out that overlap case IT-D is the network presenting the lowest standard deviation value during the CV5 process. CNN and CTT, on the other hand, did not present a sufficient increase in performance. In particular, CTT was the network that saw the smallest increase in performance compared to all the other networks. This indicates that CTT needs an extremely larger database than other networks in order to learn correctly. Furthermore, among the various networks, it was the one that presented the greatest overfitting problem. For this reason, by implementing Overlap 1 techniques, the ENGNet network and the two ITs were found to be the most suitable for classifying the ENG signal.

B. Focus on ENGNet & Leaterature comparison

Since ENGNet demonstrated superior performance among the analyzed networks, we conducted a further investigation. Instead of considering 4 classes, we considered 5, adding the "rest" class. The results are summarized in Fig. 9. This revealed that the classifier is able to recognize not only different activities, but also when no sensory activity is present. This allows us to further improve the quality associated with the ENGNet network.

Others study have been done to explore DB DA approach. An interesting one was made using ENG signal take from 56-contact cuff electrode was implanted in nine rats is reported in [21]. They evaluated four different classifiers on a non-overlapped dataset, achieving their best result with the Multi-Scale Convolution Block (MSCB) classifier, yielding an F1-

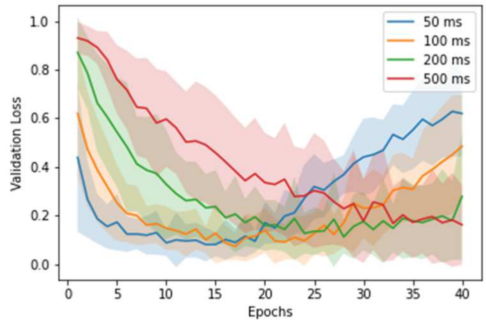


Figure 6. CNN loss value during training process.

Figure 7. CNN F1-score value obtained during DA & DB evaluation using 100ms of input window.

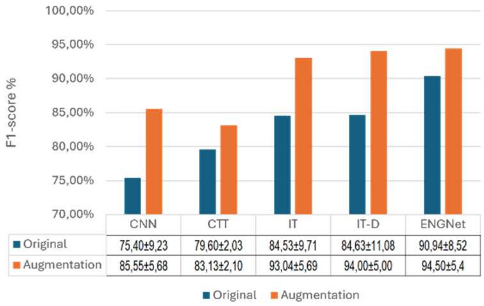
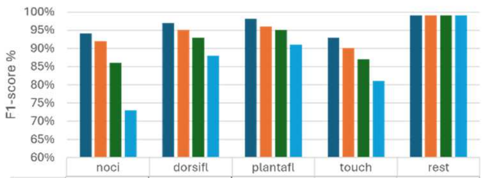


Figure 8. Neural Network F1-score value obtained using Overlap 1with 100ms of input window.



	noci	dorsifl	plantafl	touch	rest
500ms	94±2,9	97±1,2	98±1,3	93±1,5	99±0,1
200ms	92±4,7	95±0,8	96±1,8	90±6,3	99±0,1
100ms	86±4,9	93±3,7	95±5,1	87±8,6	99±0,2
50ms	73±7,9	88±5,8	91±5,3	81±9,9	99±0,3

500ms 200ms 100ms 50ms

Figure 9. ENGNet F1-score obtained using 5 class.

score of 84.6% ± 16% on a 50 ms window (1500 samples). In comparison, ENGNet demonstrated superior performance with an F1-score of 92.1% ± 6.6 across three rats, using shorter window sizes (100 ms, 500 samples). Additionally, [21] explored three data augmentation strategies, with their best approach increasing the class type with the most samples to approximately 50,000 samples per subject. This strategy resulted in an F1-score of 91.7% ± 10.3%. In contrast, ENGNet achieved an F1-score of 94.5% ± 5.4%. These findings indicate that ENGNet not only outperformed the classifiers used in paper [21] but also did so with smaller window sizes.

VI. Conclusions

In this article we address the existing gaps in ENG signal classification by taking in consideration different convolutional network. The ENGNet network, along with IT-type networks, shows promittense for managing future biomedical signal applications, especially for ENG data. Using windows equal to 100 ms were found to be the most optimal for real-time applications, reaching values up to 95% for both networks. DA strategies are able significantly to improve the results, improving F1 scores by up to 10%. In the same time subsampling stands out among DB strategies, reducing dataset size with minimal impact on performance. Overlap technique was able to improve the classification of underestimated group without compromising the prediction of the other classes. To evaluate the effectiveness in recognizing the presence or absence of a stimulus, a fifth class called "rest" was added. This class proved to be distinct and easily recognizable compared to all the others. Its inclusion provides useful information to determine whether or not activity occurs during the ENG recording.

Future research could explore IT networks using ablation studies to reduce their parameters while keeping performance unchanged or improving it. Furthermore, the use of a fifth class, capable of identifying the presence of nervous activity or not, is interesting in real-time application. Other classifiers such as Spiking Neural Networks (SNN) could be studied in the future to evaluate performance [10]. A low number of parameters, with high levels of performance and computational speed will have to be imposed to ensure that these devices can be used in implanted devices in the future.

VII. Bibliografy

[1] T. Denison and M. J. Morrell, "Neuromodulation in 2035: the neurology future forecasting series," Neurology , 2022, 98.2: 65-72.

[2] G. Hussain *et al.*, "Current status of therapeutic approaches against peripheral nerve injuries: a detailed story from injury to recovery," I*nt. J. Biol. Sci.*, 2020, 16.1: 116.

[3] K. Brattain, "Analysis of the peripheral nerve repair market in the United States," *Magellan Med Technol Consult Inc Minneap*, 2014.

[4] A. Coviello *et al.*, "Artificial neural networks-based real-time classification of ENG signals for implanted nerve interfaces," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 8, pp. 2080-2095, Aug. 2024

[5] Y. Song *et al.*, "Transformer-based spatial-temporal feature learning for EEG decoding," arXiv preprint arXiv:2106.11170 (2021).

[6] E. Brunton *et al.*, "Identification of sensory information in mixed nerves using multi-channel cuff electrodes for closed loop neural prostheses," in Proc. of Int. IEEE/EMBS Conference on Neural Engineering (NER), 2017. p. 391-394..

[7] C. Silveira *et al.*, "Influence of nerve cuff channel count and implantation site on the separability of afferent ENG," *Journal of Neural Engineering* 15.4 (2018): 046004.

[8] T. D. Farrell and R. F. Weir, "The optimal controller delay for myoelectric prostheses," in *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 1, pp. 111-118, Mar. 2007.

[9] R. Rieger *et al.*, "Design of a low-noise preamplifier for nerve cuff electrode recording," in *IEEE J. Solid-State Circuits*, vol. 38, no. 8, pp. 1373-1379, Aug. 2003.

[10] A. Coviello *et al.*, "Neural network-based classification of ENG recordings in response to naturally evoked stimulation," in Procof ACM International Conference on Nanoscale Computing and Communication (Nanocom), 2022. p. 1-7.

[11] E. Vasta, A. Coviello, U. Spagnolini, and M. Magarini, "Classification of sensory neural signals through deep learning methods," in Proc. International Conference on Smart Technologies (EUROCON), Turin, Italy, 2023, pp. 313-318.

[12] A. Vaswani *et al.*, "Attention is all you need," in *Advances in neural information processing systems,* 30 (2017).

[13] Cross-validation: Evaluating estimator performance." Accessed: 10 October 2023, [Online]. Available: https://scikit- learn.org/stable/modules/cross_validation.html#stratified-k-fold.

[14] E. Lashgari, D. Liang, and U. Maoz, "Data augmentation for deep-learning-based electroencephalography," in *J. Neurosci. Methods*, 346 (2020): 108885.

[15] H. Ali *et al.*, "Imbalance class problems in data mining: A review," *Indones. J. Electr. Eng. Comput. Sci.*, 14.3 (2019): 1560-1571.

[16] H. He and E. A. Garcia, "Learning from imbalanced data ," in *IEEE Trans. Knowl. Data Eng.* vol. 21, no. 9, pp. 1263-1284, Sept. 2009.

[17] S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng and P. J. Kennedy, "Training deep neural networks on imbalanced data sets," in Proc. of Int. Joint Conf. Neural Networks (IJCNN), 2016, pp. 4368-4374 .

[18] N. Moniz, P. Branco and L. Torgo, "Resampling strategies for imbalanced time series." In Proc. of Int. Conf. Data Science and Advanced Analytics (DSAA), 2016, pp. 282-291.

[19] Class weights for categorical loss, 2021. [Online]. Available: https://bit.ly/class-weights-categorical-loss.

[20] A. Coviello *et al.*, "Comparison of data compression methods for implanted real-time peripheral nervous system," in proc. Int. Conf. Metrology eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE), 2023, pp. 110-115.

[21] A. P. S. Gill and J. Zariffa, "Time series classification of multi-channel nerve cuff recordings using deep learning," in *Plos one* 19.3 (2024): e0299271.

[22] Williams, Ian, et al. "SenseBack-an implantable system for bidirectional neural interfacing." *IEEE Transactions on Biomedical Circuits and Systems* 14.5 (2020): 1079-1087.