Early Detection of Parkinson Disease using Deep Neural Networks on Gait Dynamics

Lerina Aversano
Dept. of Engineering
University of Sannio
Benevento, Italy
aversano@unisannio.it

Mario Luca Bernardi
Dept. of Engineering
University of Sannio
Benevento, Italy
bernardi@unisannio.it

Marta Cimitile
Unitelma Sapienza University
Rome, Italy
marta.cimitile@unitelmasapienza.it

Riccardo Pecori
Dept. of Engineering
University of Sannio
Benevento, Italy
rpecori@unisannio.it

Abstract—**Parkinson's disease is a degenerative movement disorder causing considerable disability. However, the early detection of this syndrome and of its progression rates may be decisive for the identification of appropriate therapies. For this reason, the adoption of Neural Networks to detect this disease on the base of walking information is gaining more and more interest. In this paper, we defined a Deep Neural Network based approach allowing one to exploit the information coming from various sensors located under the feet of a person. The proposed approach allows one to discriminate people affected by the Parkinson syndrome and detect the progression rates of the disease itself. To evaluate the proposed architecture we used a known dataset with the aim to compare its performance with other similar approaches. Moreover, we performed an in-depth hyper-parameter optimization to find out the best neural network configuration for the specific task. The comparison shows that the proposed classifier, trained with the best parameters, outperforms the results proviously obtained in other studies on the same dataset.**

Index Terms—**Parkinson Disease, Gait Analysis, Deep Learning, Dense Neural Networks, Parameter Optimization**

I. INTRODUCTION

Parkinson disease (PD) is a chronic and degenerative nervous system disorder that affects movements. The symptoms are an increasing difficulty to walk, to speak or to complete other simple tasks. For this reason, the monitoring of gait outcomes through wearable technology (e.g., inertial sensors) can be a useful and inexpensive alternative to evaluate the presence of this disorder in a controlled and daily living environment [10]. According to this, in the last years, a great interest was addressed towards the analysis of data extracted from sensors able to identify freeze of gait (FOG) to discriminate ill and healthy subjects [29], [9], [25], [22], [2]. However, the recognition of ill and healthy subjects when FOG already appeared is poorly useful to perform PD early identification.

Conversely, the early identification of PD and the constant monitoring of the PD severity can be very useful to identify the appropriate therapies to slow down the progression of the disease (i.e., preserving the integrity of the brain neurons) and to evaluate their effectiveness and calibrations over time. Basing on the above considerations, in this study we propose a deep learning approach allowing one to detect PD and PD severity levels by means of the analysis of data extracted from wireless sensors. Compared with traditional machine learning

techniques, the proposed deep learning approach is more suitable to handle multimodal data. Furthermore, deep learning algorithms can outperform the machine learning ones when a sufficient number of data, able to represent the complexity of the studied problem, is provided [8].

The main contribution of this paper is the application of a Deep Neural Network (DNN) architecture, to learn from data extracted by sensors located under the feet of a person to discriminate ill and healthy subjects, as well as to evaluate the progression rate of the Parkinson Disease. The effectiveness of the proposed architecture is also evaluated, for the first time in the literature, in the case of subjects that, while walking, are also involved in another activity (to perform a count down during their walking). Moreover, another contribution of this study consists in a high parameter optimization analysis aimed to find the best parameters' combination for the proposed architecture. Finally, the proposed approach is compared with similar existing approaches using the same dataset, outputting the best results ever reached in the literature. In the following Section II, a general background on Deep Learning (DL) is briefly introduced. In Section III, a brief discussion of the related work is reported. The proposed approach is described in Section IV, while the experimental results are discussed in Section V. Finally, in Section VII and VIII, the threats to validity and the conclusions are reported, respectively.

II. BACKGROUND ON DEEP LEARNING

Deep Learning is a set of recent machine learning techniques, often used in classification problems[1], allowing one to simulate the information processing of biological nervous systems [7]. A DL architecture consists of a set of connected layers. In particular, each layer extracts, from its input data, different levels of abstraction, organizing concepts in a hierarchical structure that can be used to perform feature learning and pattern classification. DL algorithms are considered more suitable (with respect to other machine learning approaches) to be applied in contexts characterized by a high level of complexity (several features and a huge number of data) in order to obtain high performance [5], [6]. Basing on the described characteristics, in the last year, several applications of DL in health informatics have been proposed with very encouraging results [16].

This work considers a multiple layers perceptron (MLP) classifier made of a feed-forward artificial neural network with at least five layers of nodes. The MLP approaches differ from the linear perceptron approaches for the presence of multiple layers and of non-linear node activation. These factors allow one to distinguish data that cannot be separated in a linear manner. Looking at the neural network training, it can be generally split into two phases: the forward and the backward phase. In the forward phase, the nodes' activation follows one another from the input layer to the output one: except for the input nodes, all the others represent neurons performing node activation through an ad-hoc function [12]. Successively, the backward phase permits one to improve the network performance by assigning to the nodes updated weights and bias values (if necessary), in order to improve the overall performance of the neural network.

III. RELATED WORK

In the recent literature, one can find various papers dealing with deep learning and gait analysis in patients affected by PD. Several recent studies have regarded the identification of freeze of gait (FOG) to distinguish ill subjects from healthy persons, as well as the degree of severity of the disease in ill patients [29], [9], [25], [22], [2]. However, FOG consists of an episodic inability to move and typically affects patients with advanced PD [29]; thus it is not useful to evaluate abnormalities occurring in the early stages of the disease. Starting from these considerations, in the last years some studies have focused on gait analysis. The gait and its related features, according to the literature, are far more suitable to identify PD in the early stages [20]. Moreover, differently from the aforementioned researches, in [17], [28] deep learning approaches are also used to evaluate the progression rates of the disease.

The contribution in [17] concerns the detection of ill and healthy people, as well as the classification of the severity of PD. According to the authors, they propose the first algorithms performing severity prediction based on a Unified PD Rating Scale. The considered dataset is the PhysioNet one[1]. The used deep learning architecture encompasses 18 one-dimensional convolutional networks, made of eight layers each, working in parallel, and then a convolutional network, composed of two fully connected layers, as well as of an output layer and of a concatenate layer, grouping together the outputs of the first 18 neural networks. The authors compare their work also with the one of Zhao at al. [31], which uses the same dataset and focuses on the same problems. However, in this contribution, the authors employ a different deep learning architecture, composed of two parallel branches, one with a 2D convolutional network and one with a recurrent neural network, namely an LSTM network. They obtained an overall accuracy of 98.61% on the whole PhysioNet dataset. The overall dataset accuracy values about severity detection have not been computed. Finally, in [28] the authors focus

on the already mentioned PhysioNet dataset and on the already investigated tasks of detecting ill and healthy persons (2 classes), as well as of multiclass severity classification. However, in this case the data from the two feet are considered separately as input of a deep neural network made of two parallel and identical branches. Both of them are constituted of a 2-layer convolutional network, followed by an attention-enhanced LSTM. The two branches are finally concatenated and submitted to a softmax layer for the final classification. The data of the dataset have been segmented according to gait cycles and, in the experiments, the authors considered both the three sub-datasets singularly and altogether. For the binary classification, the achieved accuracy is 99.07% over the overall dataset. On the other hand, for the multiclass severity classification, the accuracy result is 98.03%.

In this paper, we propose a classification approach based on a deep neural network based on perceptrons. We decided to test the proposed approach on the aforementioned PhysioNet dataset, with the aim to directly compare the obtained results with the baseline studies cited above. With respect to the cited studies, an in-depth hyper-parameter optimization phase is performed as well, in order to provide useful insights about the parameter impact over the resulting classification quality.

IV. APPROACH

In this section, we first describe the proposed feature model and then we focus on the used deep neural network architecture.

A. The proposed feature model

In this paper, the identification of PD is based on the dynamics of the Vertical Ground Reaction Force (VGRF), measured by sensors under the feet of a subject at each sampling instant. The overall set of considered features is defined in Table I. The table reports on the first column the feature acronym. For each feature, in the second column, a brief description is reported. In detail, the first eight features describe the value of vertical reaction force (in Newton), captured by sensors located in different points of the right foot. Similarly, the second group of eight features describes the value of vertical reaction force (in Newton), captured by sensors positioned in different points of the left foot. The RF total and the LF total features represent the total force under the right and left foot, respectively. Finally, some of the involved patients performed a countdown (subtracting 7 downward to zero) during their walking. According to this, we added a new binary feature (called 7Count), assuming the following values: 1 when during walking the subject performed the aforementioned countdown, or 0 in case of normal walking (no countdown is performed during the walking). In this way, we wanted to test the relevance or not of adding a further task during walking in the PD classification. This has not been performed before in the related work using this same dataset.

B. Deep Neural Network model

In this work, we employed two deep neural networks with the aim to: i) distinguish ill subjects from healthy subjects
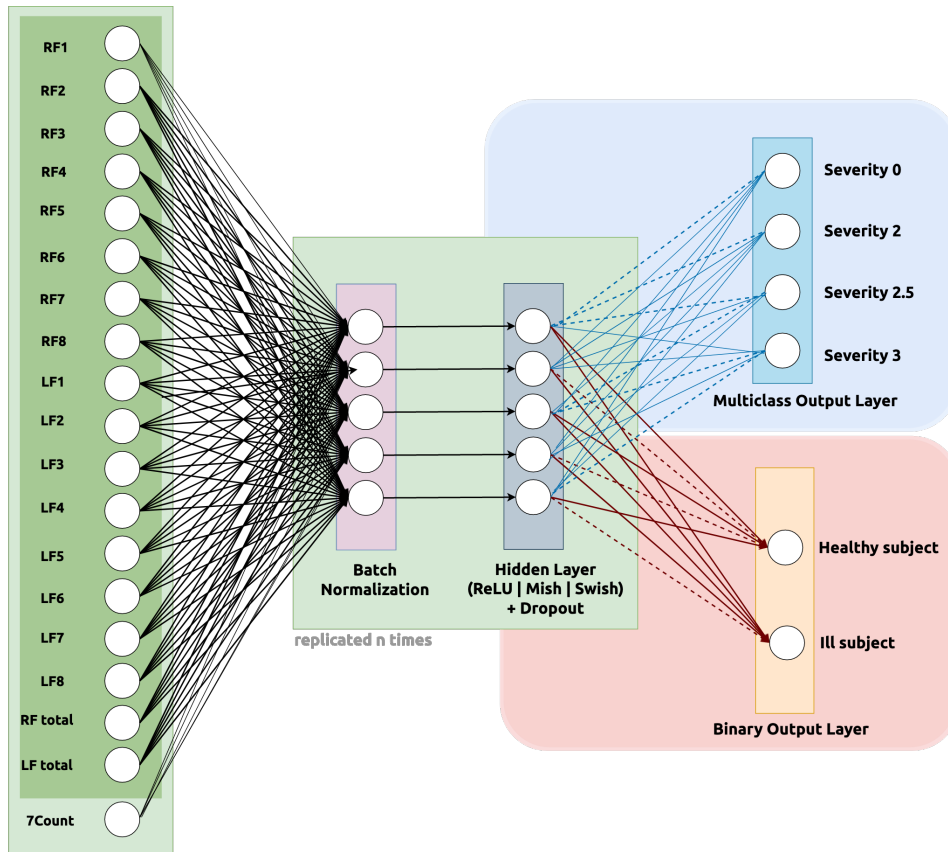
Fig. 1. The used neural network model for the binary (pink box) and multi- (blue box) classification problem.

(binary classification), ii) distinguish subjects on the base of different degrees of illness (multinomial classification). The architecture of both deep neural networks is depicted in Figure 1, consisting of a variable number of hidden layers (from six to nine) and of different output layers for the two cases. In particular, the figure shows the case of the binary output layer (in pink) allowing for the discrimination of ill subjects from healthy ones. Alternatively, the same architecture (with a different output layer) can be applied to discriminate subjects on the base of the degree of severity of the disease (the blue rectangle shows the multiclass output layer consisting of an ordinal scale of severity levels in the set $\{0, 2, 2.5, 3\}$). The overall architecture is composed of:

- one *Input layer*: the entry point of the network, encompassing a number of nodes equal to the number of considered features (the 19 of the considered feature model);
- an initial *Batch Normalization layer*: this serves to improve the training of deep feed-forward neural networks. It allows one to increase the speed of training, to adopt higher learning rates, to initialize flexible parameters in a more flexible way, as well as to saturate possible non-linearities. We added this layer because batch normalized models can provide higher accuracy on both validation and test, thanks to a stable gradient propagation within

the network itself [14].
- a variable number of *Hidden layers*: made of artificial perceptrons, whose output is calculated as a weighted sum of their inputs and passed through a certain activation function. In the evaluation section, we experimented with a different number of hidden layers, in order to reach the best performance.
- a *Dropout layer*: it immediately follows each hidden layer. This layer aims to prevent over-fitting by implementing a regularization technique. This is achieved by turning off randomly several neurons in a layer according to a certain probability p from a Bernoulli distribution. This probability is usually the same for each node of the coupled hidden layer and it usually ranges from 0.0 to 0.5.
- one *Output layer*: this layer produces the final classification outcome and is usually made of a number of neurons equal to the number of classes. We used a dense layer and a $softmax$ function for the neurons of this layer for both binary and multinomial classification problems.

The just described deep neural network model was trained by using categorical cross-entropy [18] as a loss function.

V. Experiment Description

In this section, we present the application of the deep neural network architecture described in Section IV-B on an

TABLE I
DESCRIPTION OF THE CONSIDERED FEATURES.

Acronym	Description
RF1	Vertical reaction force from the sensor located in the heel under right foot
RF2	Vertical reaction force from the sensor located in the left rear part of right foot
RF3	Vertical reaction force from the sensor located in the right rear part of right foot
RF4	Vertical reaction force from the sensor located in the left part of the inset of right foot
RF5	Vertical reaction force from the sensor located in the right part of the inset of right foot
RF6	Vertical reaction force from the sensor located in the left part of the sole of right foot
RF7	Vertical reaction force from the sensor located under the ball of right foot
RF8	Vertical reaction force from the sensor located under the toes of right foot
LF1	Vertical reaction force from the sensor located in the heel under left foot
LF2	Vertical reaction force from the sensor located in the left rear part of left foot
LF3	Vertical reaction force from the sensor located in the right rear part of left foot
LF4	Vertical reaction force from the sensor located in the left part of the inset of left foot
LF5	Vertical reaction force from the sensor located in the right part of the inset of left foot
LF6	Vertical reaction force from the sensor located in the left part of the sole of left foot
LF7	Vertical reaction force from the sensor located under the ball of left foot
LF8	Vertical reaction force from the sensor located under the toes of left foot
RF Total	Total force under right foot
LF Total	Total force under left foot
7Count	Whether the subject is counting down

OpenData dataset. In the following subsections, a description of the analyzed datasets as well as of the experiment settings are reported.

A. Dataset description

In this study, we adopted an open data dataset[2]. We decided to use this dataset for two main reasons. Firstly, the dataset is made of three different sub-datasets, each one coming from the contribution of three different neuroscience research experiments [30], [13], [11]. This ensures that the extracted measures are considered valuable in the medical community. Secondly, the only existing approaches that are comparable with the proposed one are tested on the same dataset.

The whole dataset encompasses 93 patients with idiopathic PD (59 males and 34 females) and 73 healthy control subjects (40 males and 32 females). Every participant walked in his/her usual pace for about 2 minutes, while wearing a pair of shoes with force sensors. All of these studies collected the data from 16 sensors located under the sole of each foot, 8

per foot and all the sub-datasets are consistent and contain data that can be related to the feature model proposed in Section IV-A, except for the 7Count feature (available only for Ga sub-dataset). Table II reports some statistics for the considered whole dataset (last row) and the composing sub-datasets (called Ga, Ju, and Si, respectively). For each dataset the number of considered subjects (second column) and the number of total instances (third column) are reported. The total number of subjects is then split, for each considered dataset, in four different groups representing a different level of disease severity. The referring severity scale is the Hoehn and Yahr scale[3]. It usually comprises 5 levels of severity of PD, ranging from 1 to 5. In our experiments, we considered only stages 0 (healthy subject), 2, 2.5 and 3, since these are the only stages exhibited by the patients in the considered datasets.

B. Experiment setting

Two different experiments have been carried out with the aim to evaluate the capability of our proposed classifier to distinguish, respectively: i) ill subjects and not ill subjects and ii) the level of severity of the subject's disease. Each experiment has been performed on all the datasets listed in Table II using as feature model the one described in Table I (except for the 7Count feature). Moreover, for the Ga sub-dataset, we also considered the 7Count feature and we performed a further experiment aiming to evaluate whether the value of this feature affects the obtained results. Furthermore, we performed a thorough hyper-parameter optimization step [3] to find the best combination of the parameters reported in III. The best hyper-parameters have been found using a Sequential Bayesian Model-based Optimization (SBMO) approach, implemented using the Tree Parzen Estimator (TPE) algorithm as defined in [4].

As the table summarizes, the following ranges were considered:

- **Network size**: we considered three levels of network sizes (small, medium and large), depending on the actual number of layers. A small sized network contains a maximum of 1.5 mln of learning parameters. A medium one is composed of a number of parameters between 1.5 mln and 7 mln, whereas a large network is made up of more than 7 mln and up to 22 mln parameters;
- **Activation function**: we used the well known and widely adopted ReLU activation function and we also experimented two activations function that have shown good results in recent studies, called Swish and Mish [21], [19], respectively. It is well known that ReLU suffers from the "dead" units problem: during training some ReLU units always output the same value for any input. This happens by learning a large negative bias term for its weights during training and also means that it takes no role in discriminating between inputs. When a ReLU unit ends up in this state, it is very unlikely to be subsequently recovered (because the function gradient at 0 is still 0

---

[2]https://physionet.org/content/gaitpdb/1.0.0/

[3]https://parkinsonsdisease.net/diagnosis/rating-scales-staging/

Dataset	Total Subjects	Total Instances	Severity 0	Severity 2	Severity 2.5	Severity 3	Total Patients
Ga [30]	47	1,361,382	18	15	8	6	29
Ju [13]	55	1,180,552	26	12	13	4	29
Si [11]	64	775,616	29	29	6	0	35
Whole	166	3,317,550	73	56	27	10	93

TABLE III
OPTIMIZED HYPER-PARAMETERS AND CONSIDERED RANGES.

Hyperparameters	Ranges
Batch Size	{128, 256, 512}
Network Size	{Small, Medium, Large}
Activation Functions	{ReLU, Swish, Mish}
Dropout	in range [0.1, 0.2]
Optimization algorithm	{SGD, Adam, RMSProp, Nadam, Adamax, Adagrad }
Learning Rate	in range [5, 15] (normalized, refer to text)

TABLE IV
ACCURACY OBTAINED BY THE PROPOSED BINARY CLASSIFIER COMPARED
WITH OTHER EXISTING SOLUTIONS.

Dataset	Proposed Approach	Approach 1 [17]	Approach 2 [31]	Approach 3 [28]
Ga	**99.52%**	-	98.7%	99.31%
Ju	99.29%	-	98.41%	99.29%
Si	**99.40%**	-	98.88%	99.16%
Whole	**99.37%**	98.7%	98.61%	99.07%

meaning that SGD will not alter the weights). There are variants, like "Leaky" ReLU, with a small positive gradient for negative inputs, that are an attempt to address this issue and give a chance to recover. We choose for our comparison Swish and Mish since they both do not suffer from the dead neurons issue and deal better with the vanishing gradient problem.

- **Learning rate**: it ranged from 5 to 15, normalized with respect to the optimization algorithm. For instance, using the SGD optimizer, the range was from 0.005 to 0.15;
- **Number of layers**: the numbers of considered layers varied from 5 to 9;
- **Batch size**: batch sizes greater than 512 make the training process less stable and the final accuracy was not satisfactory hence we compared three batch sizes (128, 256, and 512);
- **Optimization algorithm**: we tested several optimization algorithms to minimize the loss, such as the Stochastic Gradient Descent (SGD) [23], Adam [15], RmsProp [27], Nadam [27], Adamax [26], and Adagrad [26] optimizers. In particular, SGD has been integrated in all experimentations with Nesterov Accelerated Gradient (NAG) correction to avoid excessive changes in the parameter space, as specified in [24];
- **Dropout rate**: we considered different dropout rates belonging to the interval [0.1, 0.2] with a step of 0.05.

Both the binary classification (ill/not ill subjects) and the multinomial classification problem (classification on the base of the level of severity of the subject's disease) were performed with a changing number of epochs to validate every single considered dataset, and then, the whole merged dataset.

Four known metrics have been used to evaluate the classification results: Accuracy and Validation Accuracy, Loss and Validation Loss. The error has been evaluated with the Mean Squared Error (MSE). The accuracy has been computed as the ratio of the sum of true positives and true negatives to the total

number of classified samples in the test set. The validation accuracy is the accuracy calculated on the validation dataset. The loss implies how poorly or well a model behaves after each iteration of optimization.

VI. RESULTS AND DISCUSSION

In this section, we only discuss some of the obtained results. Figure 2 shows the evaluation of the accuracy and the validation accuracy of the binary classifier, across a 10-fold cross-validation process and versus an increasing number of epochs, for each of the considered datasets (the three sub-datasets and the whole unified one). For the Ga and Si datasets, good results are obtained with a number of layers equal to five. For the Ju and the whole dataset, the best results are obtained when the number of layers is 8. In both cases the best results were obtained using mish activation function. It can be seen that, for each considered dataset, the validation accuracy is better than the training accuracy with a very smooth trend of the curve an with a top value about at the 100th epoch for Ga and Si and 150 epochs for Ju and the whole datasets. The best obtained accuracy values for each considered dataset are reported in the second column of Table IV. In the columns from three to five, the table shows the corresponding values obtained in other comparable studies (the reference is reported as well). Looking at the table, we can conclude that the obtained accuracy is almost always greater than the accuracy obtained using alternative approaches on the same datasets.

As concerns the multi-classifier, Figure 3 shows the accuracy and the validation accuracy, across a 10-fold cross validation process and versus an increasing number of epochs, for each of the considered datasets. The number of layers providing the best results for the used DNN is 5 for datasets Ga, Ju, and Si, while it is 7 for the whole merged dataset. The same considerations drawn in the previous paragraph hold: for each considered dataset, the validation accuracy increases smoothly and is quite better than the training accuracy. The best values are reached about at the 100th epoch for the three sub-datasets and at the 200th epoch for the merged dataset.
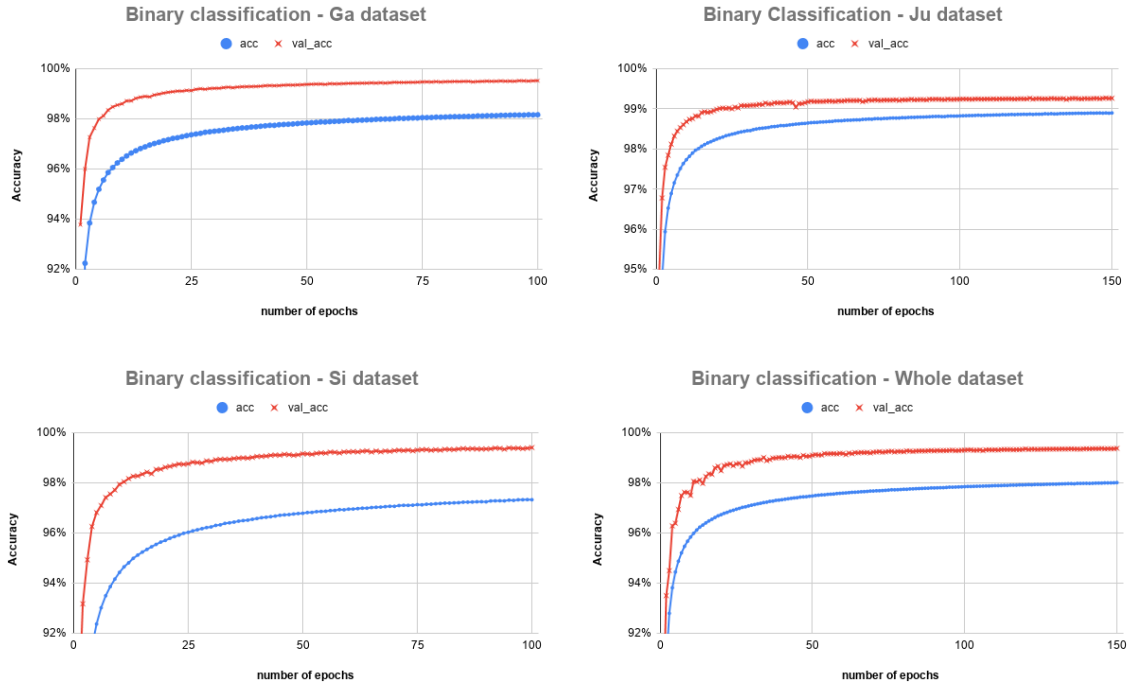
Fig. 2. Accuracy and validation accuracy, for each considered dataset, using the binary classifier.
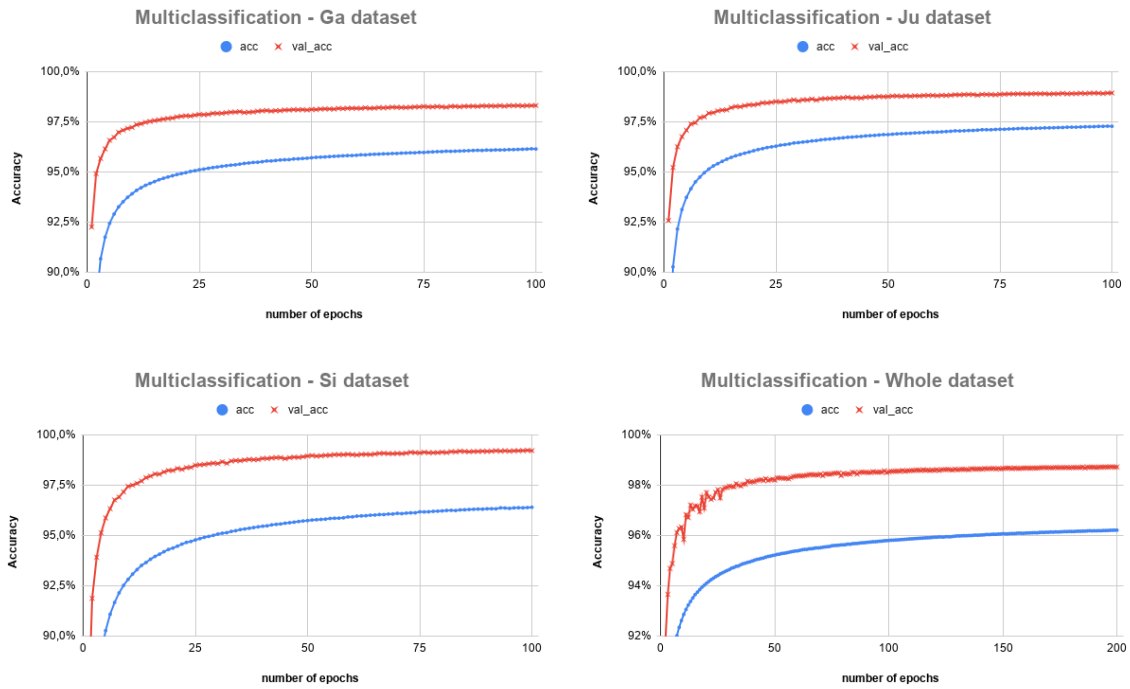

Fig. 3. Accuracy and validation accuracy, for each considered dataset, obtained by using the multi-classifier.

TABLE V

THE LIST OF THE 6 PERMUTATIONS PROVIDING THE BEST VALIDATION ACCURACY ON THE WHOLE DATASET FOR THE MULTINOMIAL CLASSIFICATION.

ID	Network Size	Activation Function	Learning Rate	No. Layers	Batch size	Optimization Algorithm	Dropout Rate	Val. Acc.	Val. Loss	Training Time (sec)	No. Param. (mln)
1	medium	mish	12	8	128	SGD	0.20	**0.9910**	0.019	34604.15	12
2	medium	swish	8	8	128	SGD	0.15	0.9899	0.019	24481.79	12
3	medium	relu	12	8	128	RMSProp	0.20	0.9885	0.023	34375.59	12
4	small	relu	9	7	256	SGD	0.15	0.9881	0.024	15685.67	7
5	small	relu	9	7	256	Nadam	0.15	0.9877	0.026	18881.39	7
6	small	relu	9	6	256	SGD	0.15	0.9857	0.032	15595.17	0.69

Moreover, Figure 4 shows, for the binary classifier (left side) and for the multi-classifier (right side), the values obtained for the accuracy and validation accuracy when the $7Count$ feature is considered or not, respectively. The adopted dataset for this experiment is Ga, since it is the only one containing also this further feature. Looking at the figure, we can observe that, when the 7Count feature is used, the obtained results are a little improved of about 0.2%-0.3%.

Finally, Table V reports the list of the parameters sets giving the best validation accuracies in the multi-classification experiment performed on the whole dataset. In the table, we report the validation accuracy, the validation loss, the training time and the number of parameters (mln) obtained for different combinations of: Network Size, activation function, learning rate, number of layers, batch size, optimization algorithm and dropout rate. As shown in the table, the best permutation has a dropout rate equal to 0.20, while almost all the remaining have a value equal to 0.15. It is also interesting to observe that, among smaller networks, ReLU is still competitive compared with Mish and Swish. For medium and large networks Swish and Mish perform better instead. This is also confirmed by a direct comparison among different activation functions reported in Figure 5, showing the trend of accuracy for ReLu, Mish and Swish for some of the obtained trials of networks with different numbers of layers. As shown, for seven layers networks, Swish and Mish are almost equivalent while Mish on 8 layers is quite more stable and performs better. This is also confirmed by the best validation accuracy we obtained (i.e., 0.991): it holds for a medium 8-layers network with a slightly higher learning rate and a Mish activation function.

VII. THREATS TO VALIDITY

Looking at the construct validity threats, some imprecisions and omissions can be due to the sensors used to extract the considered features. To avoid this limitation, we considered three different datasets using different sensors to extract the same measures.

Moreover, looking at the internal validity, if the adopted datasets are not correctly labeled or are obtained with a non-rigorous process, we will have classification errors. This risk is strongly mitigated because the used datasets are well documented and referenced in medical studies.

Finally, threats to external validity concern the generalization of the discussed findings. We have evaluated our approach on a relevant number of subjects coming from three existing datasets having different size, characteristics, and previously

adopted with different goals. In any case, in the future, it is possible to further analyze more datasets with more subjects.

VIII. CONCLUSIONS AND FUTURE WORK

In this paper, a Deep Learning architecture is proposed to exploit information coming from a set of sensors located under the feet of a person. This information is used to discriminate persons affected by PD and to identify the severity of their disease. Moreover, in this study, we have performed a great parameter optimization to evaluate, for the proposed classifiers, the best parameters, basing on the obtained accuracy. The approaches have been tested on three known datasets and the classification is performed considering both two and four classes. The obtained results show for all considered datasets very good results, obtaining (in the best case) a validation accuracy of 0.991 for the identification of the disease severity. Generally, the obtained results are better if compared with the results obtained on the same datasets using similar approaches presented in the literature. As future work, we will extend the high parameter optimization step and we aim to augment the considered set of features. Finally, further experimentation will be performed to generalize the obtained results to a multi-classification with much more degrees of severity.

REFERENCES

[1] P. Addabbo, M. Focareta, S. Marcuccio, C. Votto, and S.L. Ullo. Land cover classification and monitoring through multisensor image and data combination. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 902–905, 2016.

[2] J. Ajay, C. Song, A. Wang, J. Langan, Z. Li, and W. Xu. A pervasive and sensor-free deep learning system for parkinsonian gait analysis. In *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, pages 108–111, March 2018.

[3] Yoshua Bengio. Gradient-based optimization of hyperparameters. *Neural Computation*, 12(8):1889–1900, 2000.

[4] James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for hyper-parameter optimization. In *Proceedings of the 24th International Conference on Neural Information Processing Systems*, NIPS'11, pages 2546–2554, USA, 2011. Curran Associates Inc.

[5] Mario Luca Bernardi, Marta Cimitile, Fabio Martinelli, and Francesco Mercaldo. Keystroke analysis for user identification using deep neural networks. In *International Joint Conference on Neural Networks, IJCNN 2019 Budapest, Hungary, July 14-19, 2019*, pages 1–8. IEEE, 2019.

[6] M.L. Bernardi, M. Cimitile, F. Martinelli, and F. Mercaldo. Driver and path detection through time-series classification. *Journal of Advanced Transportation*, 2018, 2018.

[7] Li Deng, Dong Yu, et al. Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, 7(3–4):197–387, 2014.

[8] Julià Camps et al. Deep Learning for Detecting Freezing of Gait Episodes in Parkinson's Disease Based on Accelerometers. In Ignacio Rojas, Gonzalo Joya, and Andreu Catala, editors, *Advances in Computational Intelligence*, pages 344–355, Cham, 2017. Springer International Publishing.
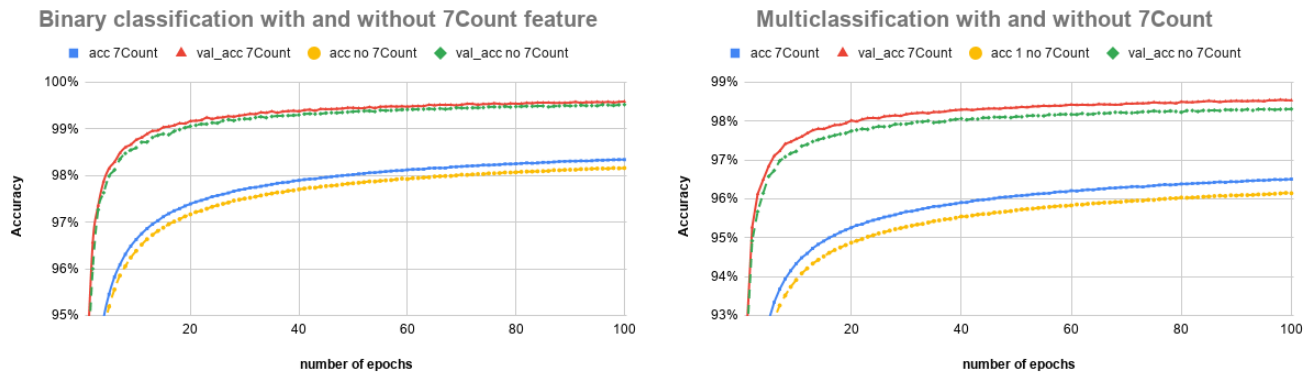
Fig. 4. Accuracy obtained considering or not the 7count feature in the binary and in the multi-classification.
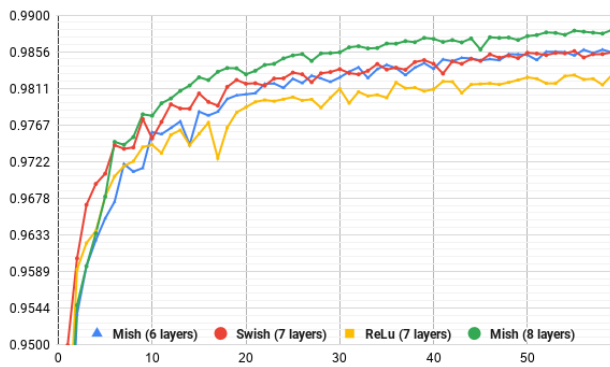

Fig. 5. Accuracy using different activation functions and numbers of layers versus the number of epochs.

[9] Julià Camps et al. Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit. *Knowledge-Based Systems*, 139:119 – 131, 2018.

[10] Matteo Gadaleta et al. Deep learning techniques for improving digital gait segmentation. *arXiv*, 1907.04281, 2019.

[11] Silvi Frenkel-Toledo, Nir Giladi, Chava Peretz, Talia Herman, Leor Gruendlinger, and Jeffrey M. Hausdorff. Treadmill walking as an external pacemaker to improve gait rhythm and stability in parkinson's disease. *Movement Disorders*, 20(9):1109–1114, 2005.

[12] Matt W Gardner and SR Dorling. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment*, 32(14):2627–2636, 1998.

[13] Jeffrey M. Hausdorff, Justine Lowenthal, Talia Herman, Leor Gruendlinger, Chava Peretz, and Nir Giladi. Rhythmic auditory stimulation modulates gait variability in parkinson's disease. *European Journal of Neuroscience*, 26(8):2369–2375, 2007.

[14] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pages 448–456. JMLR.org, 2015.

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

[16] Gloria Hyun-Jung Kwak and Pan Hui. Deephealth: Deep learning for health informatics. *ArXiv*, abs/1909.00384, 2019.

[17] Imanne El Maachi, Guillaume-Alexandre Bilodeau, and Wassim Bouachir. Deep 1d-convnet for accurate parkinson disease detection and severity prediction from gait. *arXiv*, 1910.11509, 2019.

[18] Shie Mannor, Dori Peleg, and Reuven Rubinstein. The cross entropy method for classification. In *Proceedings of the 22Nd International Conference on Machine Learning*, ICML '05, pages 561–568, New York, NY, USA, 2005. ACM.

[19] Diganta Misra. Mish: A self regularized non-monotonic neural activation function. *arXiv*, 1908.08681, 2019.

[20] Sanson F et al. Pistacchi M, Gioulis M. Gait analysis and clinical correlations in early parkinson's disease. *Funct Neurol.*, 32, 2017.

[21] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. Searching for activation functions. *arXiv*, abs/1710.05941, 2017.

[22] R. Ramakrishnan, M. S. Ram, P. Pabitha, and R. Shenbaga Moorthy. Freezing of gait prediction in parkinsons patients using neural network. In *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, pages 61–66, June 2018.

[23] Tom Schaul, Ioannis Antonoglou, and David Silver. Unit tests for stochastic optimization. *arXiv*, abs/1312.6055, 2013.

[24] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ICML'13, pages III–1139–III–1147. JMLR.org, 2013.

[25] V. G. Torvi, A. Bhattacharya, and S. Chakraborty. Deep domain adaptation to predict freezing of gait in patients with parkinson's disease. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1001–1006, Dec 2018.

[26] S. Vani and T. V. M. Rao. An experimental approach towards the performance assessment of various optimizers on convolutional neural network. In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pages 331–336, April 2019.

[27] Y. Wang, J. Liu, J. Mišić, V. B. Mišić, S. Lv, and X. Chang. Assessing optimizer impact on dnn model sensitivity to adversarial examples. *IEEE Access*, 7:152766–152776, 2019.

[28] Y. Xia, Z. Yao, Q. Ye, and N. Cheng. A dual-modal attention-enhanced deep learning network for quantification of parkinson's disease characteristics. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, pages 1–1, 2019.

[29] Yi Xia, Jun Zhang, Qiang Ye, Nan Cheng, Yixiang Lu, and Dexiang Zhang. Evaluation of deep convolutional neural networks for detection of freezing of gait in parkinson's disease patients. *Biomedical Signal Processing and Control*, 46:221 – 230, 2018.

[30] Galit Yogev, Nir Giladi, Chava Peretz, Shmuel Springer, Ely S. Simon, and Jeffrey M. Hausdorff. Dual tasking, gait rhythmicity, and parkinson's disease: Which aspects of gait are attention demanding? *European Journal of Neuroscience*, 22(5):1248–1256, 2005.

[31] Aite Zhao, Lin Qi, Jie Li, Junyu Dong, and Hui Yu. A hybrid spatio-temporal model for detection and severity rating of parkinson's disease from gait data. *Neurocomputing*, 315:1 – 8, 2018.