Fuzzy Neural Networks to Detect Parkinson Disease

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*—In this paper, we present a Deep Learning architecture, exploiting a fuzzy layer, applied to the data coming from various sensors located under the feet of a patient affected by the Parkinson's disease. The solution we propose permits one to cluster data coming from different sensors into different fuzzy partitions, according to the different parts of the feet, and to discriminate the illness of a person as well as the severity degree of the disease itself. We employed a known dataset to evaluate our solution and compared its performance with some similar approaches found in the relevant literature. Moreover, we performed an intensive parameter optimization step to find the best setting for the proposed fuzzy neural network. The evaluation shows that our solution obtains good classification results both in the binary and in the multiclassification approach.**

***Index Terms*—Parkinson Disease, Deep Learning, Fuzzy Neural Networks, Unsupervised pre-training**

I. Introduction

Parkinson's disease (PD) is a chronic and degenerative illness affecting the nervous system, mainly impairing the capabilities to move. As a matter of fact, the disease produces an increasing difficulty to walk, to utter words or to accomplish other simple daily actions.

Recent studies [1] demonstrated that monitoring the gait of people thorough wearable technology (e.g., inertial sensors) can be a useful and inexpensive solution to evaluate the presence of the PD. Therefore, in the last years, many researches [2], [3], [4], [5], [6] have been carried out to analyze data coming from inertial sensors, with the aim to identify the so-called "freeze of gait" (FOG) or to discriminate ill and healthy subjects. However, FOG detection is poorly useful to perform PD early identification, since FOG takes place when the disease has already reached a certain degree. Conversely, the early identification of PD and the constant monitoring of its severity can result into a powerful means to find out appropriate therapies, in time to slow down the progression of the disease, to save the integrity of the neurons, and to evaluate the effectiveness and the adjustments of the treatments themselves.

Given the aforementioned considerations, in this study we propose a deep learning approach in order to distinguish between ill and healthy patients, as well as the PD severity levels, by analyzing data coming from wireless sensors measuring the vertical reaction force of different parts of the feet of a subject. In the broad context of machine learning, the proposed deep learning approach is more suitable to handle multimodal and noisy data and can reveal useful information about the parts of a foot more important to analyze. Furthermore, the clustering performed by the fuzzy layer is a sort of unsupervised pre-training, which has already proved to help deep learning architectures by guiding the learning process toward basins of attraction of minima supporting better generalization from the training dataset [7]. Finally, like all deep learning algorithms, the proposed solution can outperform traditional machine learning techniques when a sufficient number of data is available [8].

Summarizing, the main contributions of this paper are:

- the definition of a Fuzzy Neural Network (FNN) architecture to analyze data coming from sensors located under the feet of a person, as well as to perform both a binary classification (ill and healthy subjects) and a multiclassification (different severity degrees of the PD);
- the evaluation, for the first time in the literature, of a feature that takes into account whether a patient is walking and simultaneously counting down by continuously subtracting a given amount;
- a great optimization campaign of the parameters of the FNN, aimed at finding the best combination for the proposed architecture;
- a comparison of the proposed architecture with similar neural networks approaches, using the same dataset, discussing the trade-off between accuracy, smaller complexity of the model and advantages of the initial fuzzy clustering.

The rest of the paper is structured as follows. Some backgrounds on Deep Learning and Fuzzy Neural Networks are introduced in Section II. In Section III, a brief discussion about the application of Deep Learning to investigate the PD is reported. The proposed architecture and feature model are carefully described in Section IV, while the experimental outcomes are presented and discussed in Section V. Finally, sections VII and VIII contain the threats to validity and the conclusions, respectively.

II. Background

A. Deep Learning

Deep Learning (DL) is a group of renowned machine learning techniques, based on artificial neural networks, that allows one to simulate the information processing of biological

978-1-7281-6932-3/20/$31.00 ©2020 IEEE

nervous systems, made of various perceptrons' layers [9]. Artificial neural networks have been devised in the past century, but they have recently come back to the attention of the research community thanks to the developments in the computational power of computers, fostering the adoption of DL architectures, made of several related layers, each one composed, in turn, of hundreds or thousands of neurons.

More in detail, each layer receives input data and abstracts and organize them into a sort of hierarchy, useful to learn features as well as to classify different patterns. Compared to traditional machine learning techniques, DL algorithms are considered i) much more suitable in contexts featuring a high level of complexity, i.e., characterized by several features and a huge number of data, and ii) capable of obtaining very high performance. Thanks to the aforementioned characteristics, in the last yeas, several applications of DL took place in different fields and health informatics is one of these, where DL has achieved very encouraging results [10].

Looking at the neural network training, this can be usually split into two main phases: the forward and the backward phase. In the former, the activation of the internal nodes, representing neurons or perceptrons, follows a certain activation function and is performed, layer after layer, from the input of the network to the output [11]. Conversely, the latter phase allows the improvement of the network performance by assigning to the nodes updated weights and bias values, if necessary.

As regards this specific paper, we consider a multiple feed-forward artificial neural network composed of a varying number of layers, each one made up of a varying number of perceptrons. The considered approach differs from the linear perceptron approaches because of the multiple layers and the non-linear node activation functions we used. These characteristics allowed us to discriminate among data that cannot be separated in a linear way, such as the ones coming from the considered sensors.

B. Fuzzy Neural Networks

Fuzzy logic tackles uncertainty, always present in real-world experiments, considering the membership degree of a feature to a certain set as a continuous function. Such a set is called a fuzzy set, opposite to a crisp set, where the membership degree is strictly complete (value equal to 1) or null (value equal to 0).

A fuzzy neural network is a particular type of neural network which is combined in a particular way with a system based on fuzzy logic [12], [13].
Neural networks and fuzzy systems share a common characteristic: they can be both used to face problems that cannot be described by any exact mathematical model. Moreover, the disadvantages they have on their own almost completely disappear using them in combination.

On one side, neural networks can only be used successfully if a sufficient amount of observed samples is available and it is not always straightforward to extract comprehensible rules or information from the structure of the neural network itself. However, no prior knowledge about the problem needs to be given.

On the other hand, a fuzzy system requires in advance linguistic rules as prior knowledge. If this knowledge is incomplete, wrong or contradictory, then the fuzzy system must be tuned, usually in a heuristic way. However, fuzzy systems are simpler to be implemented and interpreted.

Despite the different approaches to model a fuzzy neural network [14], most of them are based on some common features such as: i) the training is provided by a data-driven learning method derived from neural network theory, ii) the learning procedure is constrained to ensure the semantic properties of the underlying fuzzy system at any time of the learning process, iii) the fuzzy layer can be considered as a particular three-layer feed-forward neural network, whose second layer symbolizes the fuzzy rules.

Approximately, there can be three different kinds of fuzzy neural networks [15]:

- cooperative, where an artificial neural network and a fuzzy system work independently;
- concurrent, where their working is mutually dependent, but the neural network and the fuzzy system are still separated entities;
- hybrid, where the neural network and the fuzzy system are fused together and the fuzzy system can be interpreted as a special type of neural network.

In this paper, we are going to consider the third type of FNNs, wherein a fuzzy layer is inserted directly in a broader dense neural network.

III. Related Work

Recently, one can find various contributions in the literature dealing with deep learning and gait analysis of patients affected by the PD. Several recent papers have studied FOG to distinguish ill subjects from healthy people as well as to discriminate the degree of severity of the PD in ill patients [2], [3], [4], [5], [6]. However, FOG mirrors an irregular inability to move and usually regards patients affected by an advanced PD [2]. As a consequence, it is not so useful to evaluate anomalies happening in the very first phases of the disease itself.

Because of the aforementioned considerations, in the last years some studies focusing onto the direct analysis of the gait have started to appear in the literature. This type of analyses seems to be more suitable to identify the PD in the early phases [16].

Moreover, in [17] and [18] deep learning approaches, applied to gait data, have been also employed to evaluate the progression rates of the disease itself. The contribution in [17] focuses onto the identification of ill and healthy people as well as on the classification of the severity of the PD. According to the authors, they propose the first algorithms performing severity prediction on the basis of a Unified Parkinson's Disease Rating Scale. The considered dataset is the PhysioNet[1] one,

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

which has been used also in our contribution. The used deep learning architecture encompasses eighteen 1D convolutional networks, each made of eight layers, one for each input feature. These networks work in parallel and are followed by a convolutional network, constituted of 2 fully connected layers, and by an output layer. Finally, a concatenation layer groups together the outputs of the first eighteen neural networks. In the experiments, a 10-fold cross validation is employed, with 70% of ill patients and 30% of healthy subjects for each fold. The achieved results, in terms of accuracy, are 98.7% for the binary detection problem (ill patient detection), and 85.3% for the multiclass problem (severity detection).
The authors compare their work also with the one of Zhao at al. [19], analyzing the same dataset and focusing on the same problems. However, in this contribution, the authors employ a different deep learning architecture, made of 2 parallel branches, one with a 2D convolutional network and one with an LSTM network. The obtained overall accuracy achieved on the whole PhysioNet datset is 98.61% for the binary classification problem. The overall dataset accuracy values about severity detection are also relevant, ranging from 97.48% for the Ga sub-dataset, to 97.86% for the Ju sub-dataset, and to 98.8% for the Si sub-dataset. For the whole dataset, made of the merging of the three sub-datasets (Ga, Ju and Si), the accuracy has not been computed instead.

Finally, in [18] the authors concentrate again 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, like different inputs of a deep neural network made of two parallel and identical branches. Both of them are composed 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. As concerns the binary classification, the achieved accuracy values are 99.31%, 99.29%, 99.16% for the three single sub-datasets and 99.07% over the whole merged dataset. On the other hand, for the multiclass severity classification, the accuracy results are the following: 98.11%, 98.36%, and 99.01% for the three sub-datasets considered singularly, and 98.03% for the whole dataset.

We decided to test the approach we propose in this paper on the already mentioned PhysioNet dataset, with the aim to directly compare the obtained results with the contributions cited above by means of a great phase of hyper-parameters' optimization.

IV. Approach

In this section, we first describe the proposed feature model and then we detail the architecture of the used deep fuzzy neural network.

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 or not

A. The proposed feature model

In this paper, we assume that PD can be detected exploiting the dynamics of the Vertical Ground Reaction Forces (VGRFs), measured by sensors under the feet of a subject at a constant sampling rate, namely 100 Hz. Thus, differently from other related contributions in the literature, we do not make considerations on the gait cycle nor we adopt particular windowing techniques to consider only close temporal instants. We consider each sampling instant and its corresponding sensors' values independently, instead.

The overall set of considered features is presented and

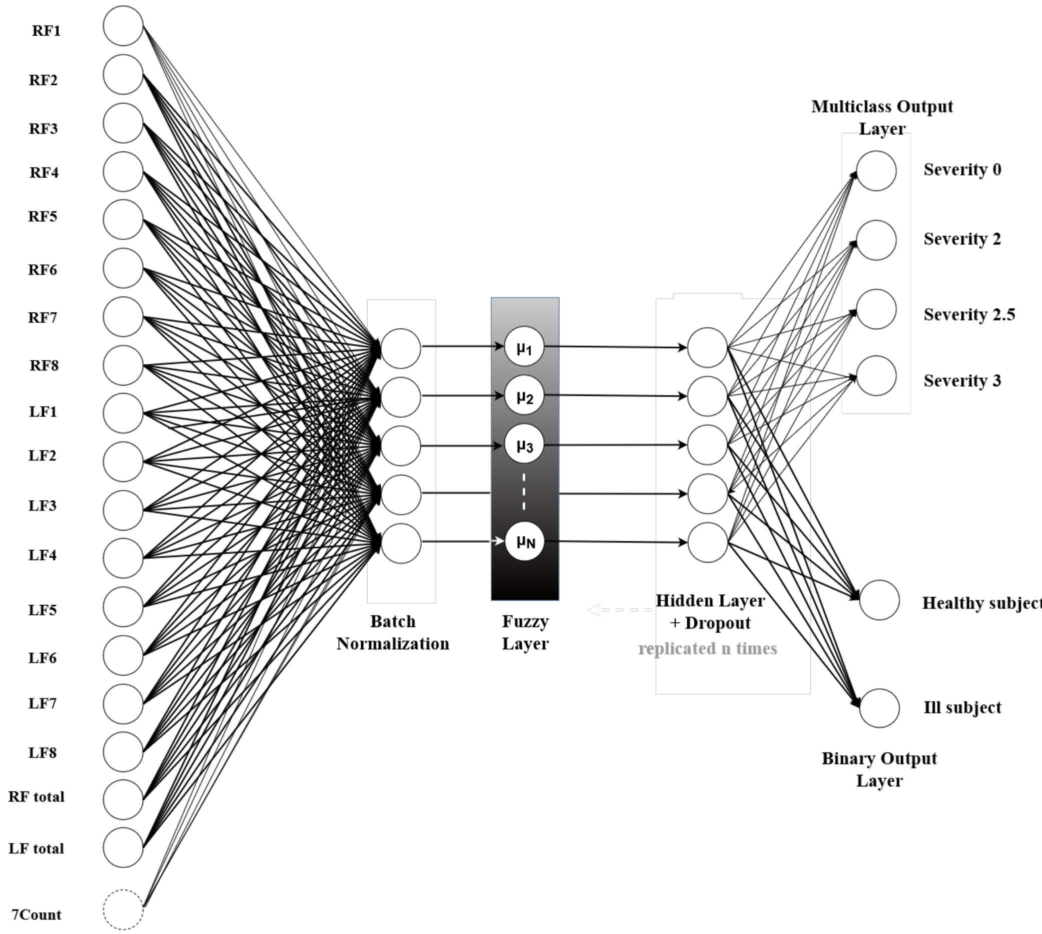

Fig. 1. The used fuzzy neural network model for the binary and multiclassification problem. μ_i represents the membership function of each fuzzy cluster, output of the Fuzzy layer.

described in Table I. The table reports, in the first column, the acronym of the feature, while, in the second column, a brief description thereof. More in detail, the first eight features describe the value of the vertical reaction force (in Newton) captured by sensors positioned in different points of the right foot. Likewise, the second group of eight features describes the value of the vertical reaction force retrieved through sensors located in different parts of the left foot. The "RF total" and the "LF total" features represent the sum of all the forces under the right and left foot, respectively.

Finally, since some of the subjects involved in the study related to Ga sub-dataset performed also a countdown (by continuously subtracting 7) while walking, we added also a further binary feature (called "7Count"), assuming the following values: 1 when during walking the subject performed the countdown, 0 in case of normal walking (no countdown). In this way, we wanted to test, for the first time in the literature, the relevance or not of this further feature on the deep learning performance.

B. Deep learning model

In this work, we employed a deep learning neural network with the aim to: i) distinguish ill subjects from healthy subjects (binary classification), ii) distinguish subjects on the base of different degrees of illness (multiclassification).

The considered architecture model is shown in Figure 1, and consists of a variable number of layers, described in the following:

- *Input layer*: the first layer of the network, receiving data from the sensors and encompassing a number of nodes equal to the number of considered features (18 or 19 in case the 7Count feature is considered);
- an initial *Batch Normalization layer*: this has been added to improve the training of the deep feed-forward neural network. Indeed, it allows also the rise in the speed of training, the adoption of higher learning rates, the initialization of parameters in a more flexible way, as well as the saturation of possible non-linearities. Moreover, this layer can furnish higher accuracy on both validation and test sets, thanks to a stable gradient propagation within the network itself [20].
- a *Fuzzy layer*: this layer is responsible for the clustering

of the initial data into interpretable groups, e.g., spatial coordinates, multi-function values, etc. The number of outputs represent the number of clusters, whose membership function is described as follows:

$$\mu_j(x, c, a) = e^{-\sum_{i=0}^{dim} 1/4 \frac{(x_i - c_i)^2}{a_i^2}}, \qquad (1)$$

where x is the input vector of dim length, c is the centroid of the j^{th} membership function and a is a vector of scaling factors. The choice of a Gaussian-like function has been made to exploit its well-known universality and smooth decreasing behavior, in this first attempt to apply fuzzy membership functions to this dataset.

- a variable number of *Hidden layers*: each of these layers contains a variable number of artificial perceptrons, whose output is a weighted sum of their inputs, passed through a certain non-linear activation function (ReLu, Swish, etc.). In the evaluation section, we will present experiments with a different number of hidden layers, as well as an optimization of the number of layers, of nodes per layer and of the used activation function, in order to test the best possible performance of the network.
- a *Dropout layer*: this layer is tightly coupled with each hidden layer that immediately precedes it. The dropout functionality has been introduced to prevent over-fitting by means of a regularization technique that turns off randomly several neurons in a hidden layer according to a probability p drawn from a Bernoulli distribution. This probability is usually chosen in the $0.0 - 0.5$ range.
- *Output layer*: this is the exit of the neural network, producing the final categorical classification and it is usually made of a number of neurons equal to the number of classes. In Figure 1, two different output layers are shown, even if one at a time is used. The binary classification (discrimination between ill and healthy subjects) is represented in the lower part, while the multiclassification problem (identification of four different degrees of severity) is in the upper part. We employed a dense layer as a final exit layer and a $softmax$ as an activation function.

V. Experiment Description

In this section, we present the application of the deep fuzzy neural network architecture described in Section IV-B on a dataset of open data. In the following subsection, a description of the analyzed dataset is provided. Moreover, the settings for the experiments we made are reported in Subsection V-B.

A. Dataset description

In this paper, we took advantage of the PhysioNet dataset[2], which we decided to use because: i) it is made of three different sub-datasets, each one coming from the contribution of three different and independent neuroscience research experiments [21], [22], [23], thus ensuring the value and validity of the extracted measures in the medical community; ii) it contains enough samples to be applied successfully in a deep learning scenario, like the one we considered in this paper.

The whole dataset includes 93 patients with idiopathic PD (59 males and 34 females) and 73 healthy control subjects (40 males and 32 females). Every participant walked with his/her usual pace for about 2 minutes, while wearing a pair of shoes with vertical reaction force sensors. All these studies collected the data from 16 sensors located under each foot, 8 per foot, and all the gathered sub-datasets are consistent and contain data that can be related to the feature model explained in Section IV-A, apart from the 7Count feature, only available for the Ga sub-dataset. Table II reports some statistics describing the considered whole dataset (last row) and the component 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 the severity of PD, where 0 is the level for healthy people. The severity scale we considered is the Hoehn and Yahr's one[3] that usually comprises 5 levels of severity of PD, ranging from 1 to 5. In our experiments, we considered only the 2, 2.5 and 3 stages, since these are the only levels 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: i) ill subjects and not ill subjects and ii) the level of severity of the disease. Each experiment has been performed on the datasets listed in Table II, using as feature model the one described in Table I. Moreover, for the Ga sub-dataset, we have also considered the 7Count feature and we have performed a further experiment to test whether and how this feature impacts the final results.

Meanwhile, we performed a thorough hyper-parameter optimization phase [24], in order to find out the best combination of the following parameters:

- Network size: we have considered in our experiments three different network sizes (small, medium and large), depending on the actual number of neurons in the various hidden layers. A small size network consists of a maximum of 1.5 mln of learning parameters. A medium size network is composed of a number of parameters between 1.5 mln and 7 mln, whereas a large network is made up of over 7 mln up to 12 mln parameters;
- Activation function: we have taken advantage of a very known activation function called ReLu and we have also tested two recent activations functions called Swish and Mish [25], respectively;
- Learning rate: we have made it vary from 5 to 15, normalized with respect to the optimization algorithm. For instance, using SGD optimizer the range is from 0.005 to 0.15;

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

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

TABLE II
STATISTICS OF THE CONSIDERED DATASETS.

Dataset	Total Subjects	Total Instances	Severity 0	Severity 2	Severity 2.5	Severity 3	Total Patients
Ga [21]	47	1,361,382	18	15	8	6	29
Ju [22]	55	1,180,552	26	12	13	4	29
Si [23]	64	775,616	29	29	6	0	35
Whole	166	3,317,550	73	56	27	10	93

- Number of layers: The numbers of considered hidden layers ranges from 6 to 9;
- Batch size: it could assume the following values: 128, 256 and 512;
- Optimization algorithm: we have tested the Stochastic Gradient Descent (SGD) [26], Adam [27], RmsProp [28], Nadam [28], Adamax [29], Adagrad [29] optimizers;
- Dropout rate: we have considered two different rates, namely 0.15 and 0.2.
- Number of outputs of the Fuzzy layer: three possible numbers of clusters have been considered, i.e., 6, 8, and 10.

Table III summarizes the considered hyper-parameters and the considered ranges or values for each hyper-parameter.

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)
Fuzzy sets	$\{6, 8, 10\}$

Both the binary classification (ill/not ill subjects) and the multiclassification problem (classification on the base of the level of severity of the subject's disease) have been performed with the fuzzy neural network, described in Section IV-B, and a changing number of epochs to validate every single considered sub-dataset, and then, the whole merged dataset.

As regards the parameters not involved in the optimization, the considered loss function has been categorical cross-entropy [30], while the chosen optimizer has been SGD, with momentum equal to 0.09 and a decay of $1e^{-6}$, accompanied with Nesterov Accelerated Gradient (NAG) correction, in order to avoid excessive changes in the parameter space, as specified in [31].

Four known metrics have been used to evaluate the classification results: accuracy and validation accuracy, loss and validation Loss. The accuracy has been evaluated over the training set and computed as the ratio of the sum of true positives and true negatives to the total number of tested instances. The validation accuracy has been computed in the same way as the the accuracy, but considering the validation set. On the other hand, the loss and validation loss, computed on the training and validation sets respectively, imply how poorly or well a model behaves after each iteration of the back-propagation mechanism.

VI. DISCUSSION OF THE RESULTS

Figure 2 shows the evaluation of the validation accuracy and validation loss of the binary classifier versus an increasing number of epochs, for sub-dataset Ga, with the parameters' configuration which permits the achievement of the best obtained results. The values of these parameters are shown in Table V, in scenario 1 row.
As one can see, the figures demonstrate that the learning process stabilizes at about 100 epochs, when both the validation accuracy and the validation loss tend to saturate. This could be considered a good value to train the considered fuzzy neural network.

In Table IV, we provide the best obtained validation accuracy values in three different scenarios with a number of epochs set to 100:

1) binary classification and Ga sub-dataset (scenario 1);
2) binary classification and whole dataset (scenario 2);
3) multiclassification and whole dataset (scenario 3).

In the columns from three to five, table IV shows the corresponding values obtained in other comparable studies, whose reference is reported as well. Looking at the table, we can see different situations: i) the achieved accuracy is between the worst and best accuracy values obtained in the literature with similar methods, which, however, do not use fuzzy layers; ii) the obtained accuracy is the worst, considering similar DL methods with no fuzzy layers at all.
These can be explained considering that the introduction of a fuzzy layer implies a discretization of the initial data, thus a sort of loss of information. However, the architecture we propose is far less complex compared with the ones in the literature and does not involve any complex consideration on gait cycle. Moreover, it results to be an intermediate solution in most of the considered scenarios and permits some further considerations that follow. As a matter of fact, looking at Table V, summarizing the best values of the validation accuracy in the three aforementioned scenarios as well as the corresponding hyper-parameters' values, we can state that: i) the accuracy decreases with the increase of the difficulty in the classification task: binary classification with a reduced dataset (Ga), then binary classification with the whole dataset, and, finally, multiclassification with the whole dataset; ii) conversely, the optimum number of fuzzy sets decrease with the difficulty of the task: this may indicate that in detecting the severity degree only some macro areas of the foot are

TABLE IV
OBTAINED ACCURACY VALUES COMPARED WITH OTHER EXISTING SOLUTIONS.

Scenario	Dataset	Proposed Approach	Approach 1 [17]	Approach 2 [19]	Approach 3 [18]
1	Ga Binary	99.10%	-	98.7%	99.31%
2	Whole Binary	95.98%	98.7%	98.61%	99.07%
3	Whole Multi	91.08%	85.3%	-	98.03%

relevant, while to detect PD in the first stages (simple binary classification) a more fine-grained distribution of the sensor data is necessary.

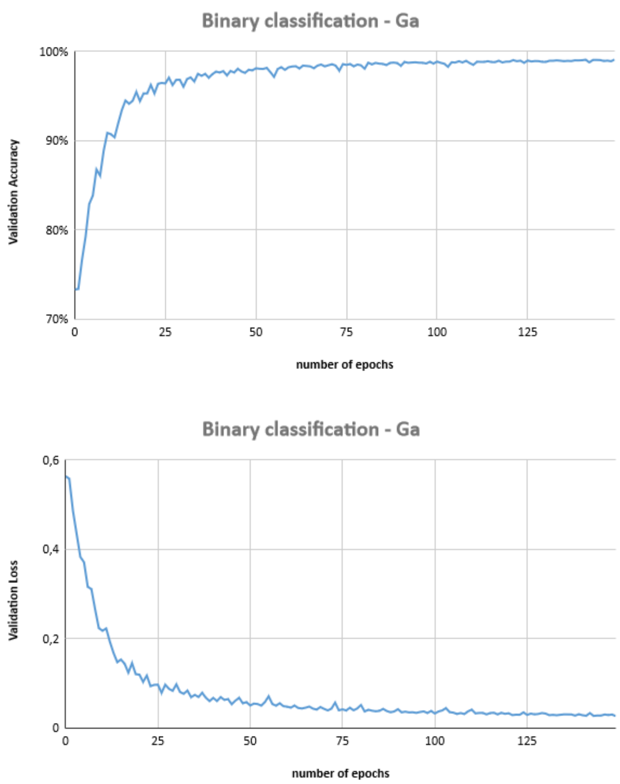

Fig. 2. Validation accuracy and validation loss trends vs the number of epochs for Ga sub-dataset in the case of binary classification.

In Fig. 3, we show the trends of the validation accuracy using or not the 7count feature, for both the binary and the multiclassification tasks. For the binary classification, the introduction of this new feature, allows the curve to be smoother from a certain number of epochs on and an increase of the performance of about 2% at the last epoch. Conversely, for the multiclassification the overall gain is much smaller in the trend, and about 1% at the last epoch. However, also in this case the optimum number of fuzzy sets is greater in the binary classification (8) than in the multiclassification (6), corroborating once again the already mentioned interpretation.

VII. THREATS TO VALIDITY

As concerns the construct validity threats, some inaccuracies and omissions can be due to the specific sensitivity of the

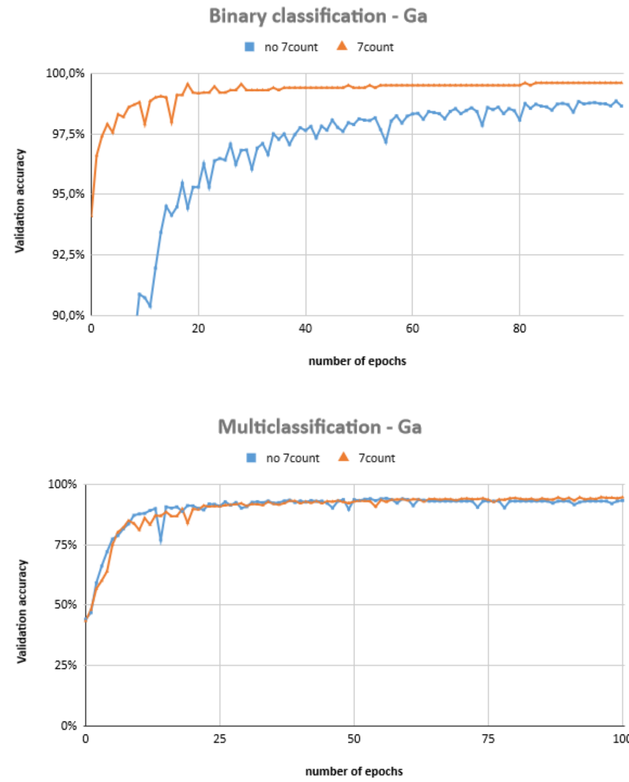

Fig. 3. Validation accuracy vs the number of epochs for Ga sub-dataset, in the case of binary and multiclassification, using or not the 7Count feature.

sensors used to extract the considered features. In order to limit this threat, we have considered three different sub-datasets, from three different and independent researches, using different sensors to extract the same measures.

Moreover, regarding the internal validity, if the adopted datasets are not correctly labeled or are obtained with a non-rigorous process, we could 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 may involve the generalization of the discussed findings. We have evaluated our approach on a great number of instances from three existing datasets having different sizes, characteristics, and previously adopted with different goals. In any case, in the future, it is possible to further analyze more datasets with more instances.

VIII. CONCLUSIONS AND FUTURE WORK

In this paper, a Fuzzy Deep Learning architecture is proposed to exploit information deriving from a set of sensors located under the feet of a person. This information has been used to discriminate persons affected by PD and to identify the severity of their disease, by grouping the data of the sensors in different fuzzy clusters.

Moreover, in this study, a high parameter optimization is performed to evaluate, for the proposed classifiers, the best possible parameters, on the basis of the obtained overall

TABLE V
THE VALUES OF THE HYPER-PARAMETERS IN THE SCENARIOS REPORTED IN TABLE IV.

Scenario	Network Size	Activation Function	Learning Rate	No. Layers	No. Fuzzy Sets	Batch size	Optimization Algorithm	Dropout Rate	Val. Acc.
1	medium	all ReLu	10	7	10	128	SGD	0.2	0.9910
2	small	all ReLu	10	6	8	128	SGD	0.2	0.9598
3	small	all ReLu	10	8	6	128	SGD	0.2	0.9108

validation accuracy. The obtained results show for all the considered datasets very good performance, obtaining (in the best case) a validation accuracy of 99.1%. The obtained results are intermediate compared with the results achieved on the same datasets using similar approaches in the literature, but our architecture is much simpler and allows for a sort of interpretability in discriminating the areas of the sole more involved in the binary or multiclassification tasks, respectively.

As future work, we will extend the high parameter optimization phase as well as the proposed set of features. Finally, further experimentation will be performed to generalize the obtained results with different and more elaborated fuzzy neural network architectures, sporting different membership functions.

REFERENCES

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

[2] Yi Xia et al. 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, and Samy Bengio. Why does unsupervised pre-training help deep learning? *J. Mach. Learn. Res.*, 11:625–660, March 2010.

[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.

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

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

[11] 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.

[12] Li-Hong Xu, You-Ling Yu, and Qi-Di Wu. General fuzzy neural network: Basic structure, algorithms and its applications. *IFAC Proceedings Volumes*, 32(2):5255 – 5260, 1999. 14th IFAC World Congress 1999, Beijing, Chia, 5-9 July.

[13] Giovanni Acampora, Mario Luca Bernardi, Marta Cimitile, Genoveffa Tortora, and Autilia Vitiello. A fuzzy clustering-based approach to study malware phylogeny. In *2018 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2018, Rio de Janeiro, Brazil, July 8-13, 2018*, pages 1–8. IEEE, 2018.

[14] Detlef Nauck and Rudolf Kruse. Neuro-fuzzy classification with nefclass. In Peter Kleinschmidt, Achim Bachem, Ulrich Derigs, Dietrich Fischer, Ulrike Leopold-Wildburger, and Rolf Möhring, editors, *Operations Research Proceedings 1995*, pages 294–299, Berlin, Heidelberg, 1996. Springer Berlin Heidelberg.

[15] Detlef Nauck, Frank Klawonn, and Rudolf Kruse. *Foundations of Neuro-Fuzzy Systems*. John Wiley & Sons, Inc., USA, 1997.

[16] Sanson F et al. Pistacchi M, Gioulis M. Gait analysis and clinical correlations in early parkinson's disease. *Functional Neurology*, 32:28–34, 2017.

[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] 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.

[19] 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.

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

[21] Galit et al. Yogev. Dual tasking, gait rhythmicity, and parkinson's disease: Which aspects of gait are attention demanding? *European Journal of Neuroscience*, 22(5):1248–1256, 2005.

[22] Jeffrey M. et al. Hausdorff. Rhythmic auditory stimulation modulates gait variability in parkinson's disease. *European Journal of Neuroscience*, 26(8):2369–2375, 2007.

[23] Silvi et al. Frenkel-Toledo. Treadmill walking as an external pacemaker to improve gait rhythm and stability in parkinson's disease. *Movement Disorders*, 20(9):1109–1114, 2005.

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

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

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

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

[28] 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.

[29] 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.

[30] 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.

[31] 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 Machine Learning - Volume 28*, ICML'13, pages III–1139–III–1147. JMLR.org, 2013.