Dipartimento di Ingegneria Elettrica e dell'Informazione "Maurizio Scarano"

Deep Transfer Learning for Alzheimer's disease detection

Nicole Dalia Cilia, Claudio De Stefano, Claudio Marrocco, Francesco Fontanella, Mario Molinara and Alessandra Scotto di Freca

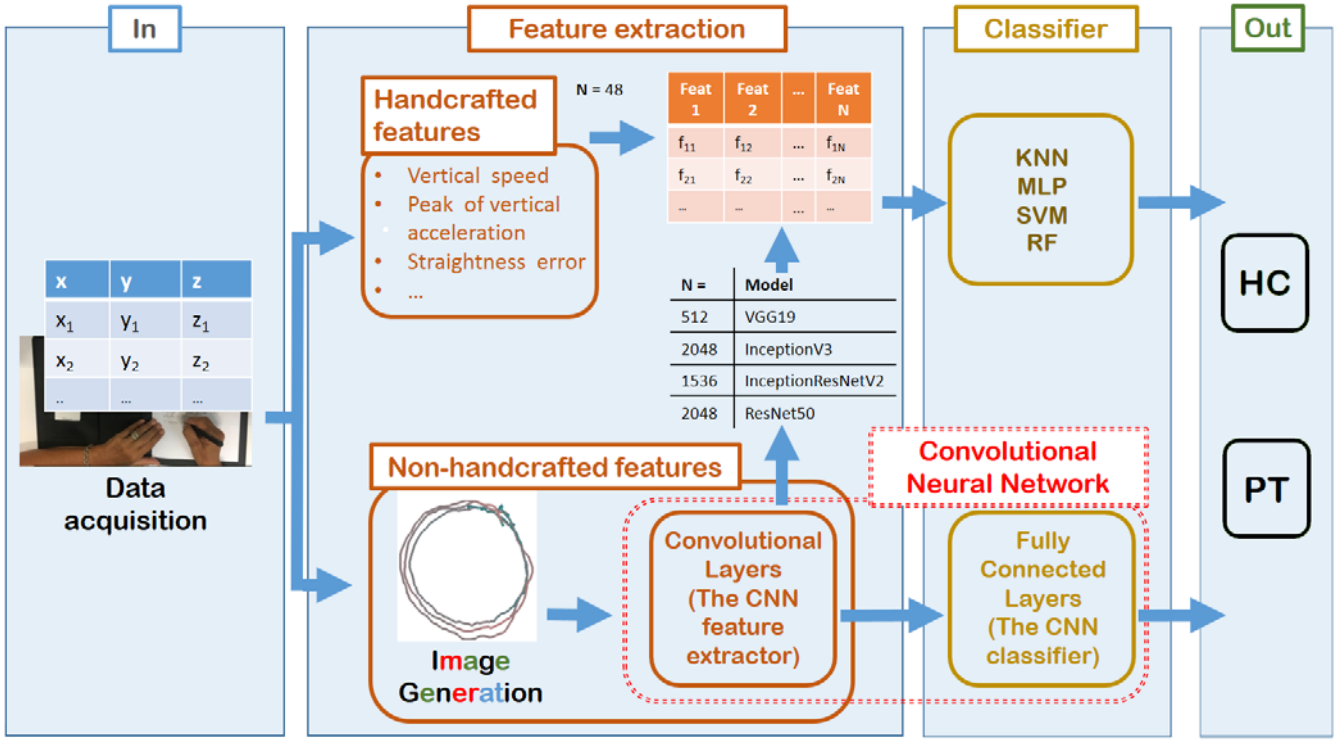
Department of Electrical and Information Engineering, University of Cassino and Southern Lazio,
Via Di Biasio 43, 03043 Cassino (FR), ITALY

Email: (nicoledalia.cilia, destefano, c.marrocco, fontanella, m.molinara, a.scotto)@unicas.it

- Early detection of Alzheimer's Disease (AD) is essential in order to initiate therapies that can reduce the effects of such a disease, improving both life quality and life expectancy of patients.
- Among all the activities carried out in our daily life, handwriting seems one of the first to be influenced by the arise of neurodegenerative diseases. For this reason, the analysis of handwriting and the study of its alterations has become of great interest in this research field in order to make a diagnosis as early as possible.
- In recent years, many studies have tried to use classification algorithms applied to handwriting to implement decision support systems for AD diagnosis. A key issue for the use of these techniques is the detection of effective features, that allow the system to distinguish the natural handwriting alterations due to age, from those caused by neurodegenerative disorders.
- In this context, many interesting results have been published in the literature in which the features have been typically selected by hand, generally considering the dynamics of the handwriting process in order to detect motor disorders closely related to AD. Features directly derived from handwriting generation models can be also very helpful for AD diagnosis.

- It should be remarked, however, that the above features do not consider changes in the shape of handwritten traces, which may occur as a consequence of neurodegenerative diseases, as well as the correlation among shape alterations and changes in the dynamics of the handwriting process.
- Moving from these considerations, the aim of this study is to verify if the combined use of both shape and dynamic features allows a decision support system to improve performance for AD diagnosis.
- To this purpose, starting from a database of on-line handwriting samples, we generated for each of them a synthetic off-line colour image, where the colour of each elementary trait encodes, in the three RGB channels, the dynamic information associated to that trait.
- Finally, we exploited the capability of Deep Neural Networks (DNN) to automatically extract features from raw images, following the Transfer Learning approach.
- The experimental comparison of the results obtained by using standard features and features extracted according the above procedure, confirmed the effectiveness of our approach.

The architecture of the whole system



the protocol consists of two tasks

- to trace a circle with a diameter of 6 cm (Task1) continuously for four times.
- to trace a circle with a diameter of 4 cm (Task2), continuously for four times.

Our dataset includes 154 subject each performing both the two tasks.

- 79 patients
- 75 healthy controls

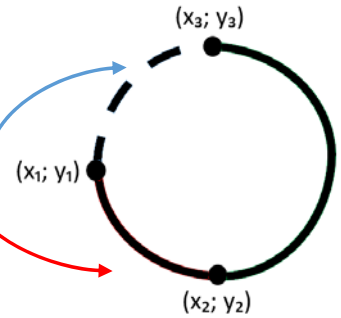
Handcrafted features → Directly derived from x, y coordinates of each acquired point together with pressure information (z coordinate)

Non handcrafted features → Derived trough the use CNNs applied to synthetic images generated from original on-line handwriting.

Handcrafted features

features are computed for each ***on paper stroke***, defined as the piece of traits between a pen down and the following pen up, or a change of direction on the y axis

features are also computed for each ***on air stroke*** , defined as the pen tip trajectories produced without touching the sheet, but keeping the pen tip close to the sheet (within the maximum allowed distance)



For each stroke we extracted 22 features grouped in two categories, namely **static** and **dynamic**.

Static features includes *Duration*, *Vertical dimension*, *Horizontal dimension*, *Inclination from the initial point to the final point, Loop surface*, *Absolute size*, etc.,

Dynamic features includes *Vertical speed peak*, *Peak of vertical acceleration*, *Jerk*, *Pen pressure* etc. (note that in case of OnAir traits only 21 features are extracted, since the pressure is equal to zero).

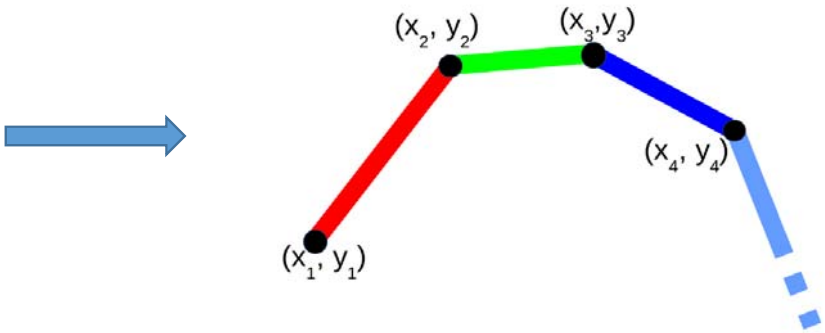
Summarizing, the **final feature vector** includes a **total number of features equal to 47** (22 features averaged over on paper traits, 21 features averaged over on air traits and 4 features relative to personal information).

Non handcrafted features

Starting from the same row coordinates used for handcrafted feature extraction, we generated the **colour images** to be submitted to the CNN networks.

For each stroke $\boldsymbol{S}_i$, we considered the following dynamic information: velocity $\boldsymbol{v}_i$, jerk $\boldsymbol{j}_i$ and pressure $\boldsymbol{z}_i$.

The values of the triplets ($\boldsymbol{z}_i$, $\boldsymbol{v}_i$, $\boldsymbol{j}_i$) have been normalized into the range [0, 255] in order to match the standard 0-255 color scale, and associated to the corresponding stroke in the generated image.



These images were used as input for the CNNs, which provided the final set of features in the *Convolutional Layer*

Non handcrafted features

- Four CNN models have been considered: VGG19, ResNet50, InceptionV3, InceptionResNetV2
- The training phase was conducted by using a validation set in order to reduce or avoid over-fitting of the network on training set.
- A 5-fold validation strategy has been considered where, for each fold used as test set (composed of 20% of images), we included 70% of images in the training set and 10% of images in the validation set.
- Note that the images included in the validation set have been randomly selected from the folds not used as test set.
- After the training phase, the CNN networks have been used on one hand for extracting non-handcrafted features (represented by the weights of the *Convolutional Layer)*, and on the other hand for classification with the final *Fully Connected Layer* (the classifier section of the deep network).

Classification

Using the two different feature extraction methodologies (hand-crafted and non-handcrafted) we obtained 5 different datasets:

- the first one consisting of feature vectors including hand-crafted features
- 4 datasets consisting of feature vectors including non-handcrafted features obtained through the four considered CNNs, respectively

For the classification step, we used four different classification schemes:

- the Random Forest (RF)
- the Multi-Layer Perceptron (MLP)
- the Support Vector Machines (SVM)
- the K-Nearest Neighbors (K-NN)

Experimental results

RESULTS OF CLASSIFICATION WITH HANDCRAFTED FEATURES

	TASK 01		TASK 02		ALL	
	ACC	SD	ACC	SD	ACC	SD
RF	60.92	3.35	68.15	2.53	68.41	1.40
K-NN	61.48	2.99	64.53	2.22	62.71	1.86
SVM	54.40	0.75	51.47	0.56	51.73	0.77
MLP	58.20	3.58	66.15	2.59	64.20	1.81

Experimental results

CLASSIFICATION RESULTS WITH NON-HANDCRAFTED FEATURES

InceptionResNetV2

	TASK 01		TASK 02		ALL	
	ACC	SD	ACC	SD	ACC	SD
RF	72.06	1.48	67.35	2.33	70.25	1.31
K-NN	66.32	1.90	66.95	2.91	64.23	1.61
SVM	70.39	1.89	67.35	1.68	69.75	0.96
MLP	66.71	3.49	56.03	2.59	55.59	1.32

InceptionV3

	TASK 01		TASK 02		ALL	
	ACC	SD	ACC	SD	ACC	SD
RF	**73.77**	2.05	**71.06**	2.08	**72.48**	1.21
K-NN	68.52	2.09	59.90	1.87	62.78	1.53
SVM	72.10	1.72	70.83	1.63	71.49	1.30
MLP	62.90	3.97	65.86	3.15	68.17	2.35

ResNet50

	TASK 01		TASK 02		ALL	
	ACC	SD	ACC	SD	ACC	SD
RF	72.65	1.03	69.11	2.94	70.28	1.47
K-NN	64.65	2.53	61.92	2.48	62.58	1.70
SVM	71.84	1.24	65.79	2.29	69.13	1.57
MLP	54.97	0.97	55.13	1.97	54.59	0.48

VGG19

	TASK 01		TASK 02		ALL	
	ACC	SD	ACC	SD	ACC	SD
RF	70.52	1.68	68.68	2.46	68.86	1.18
K-NN	66.26	2.33	64.24	2.34	64.56	1.46
SVM	70.97	1.40	66.29	1.34	68.30	0.86
MLP	64.45	2.41	59.87	2.97	58.82	2.48

Experimental results

Comparison of the best results obtained for both handcrafted and non handcrafted features.

	TASK 01					TASK 02				
	Non-Handcrafted			Handcrafted		Non-Handcrafted			Handcrafted	
	NET	ACC	SD	ACC	SD	NET	ACC	SD	ACC	SD
RF	Inc. V3	73.77	2.05	60.92	3.35	Inc. V3	71.06	2.08	68.15	2.53
K-NN	Inc. V3	68.52	2.09	61.48	2.99	Inc. Res.	66.95	2.91	64.53	2.22
SVM	VGG	70.97	1.40	54.40	0.75	Inc. V3	70.83	1.63	51.47	0.56
MLP	Inc. Res.	66.71	3.48	58.2	3.57	Inc. V3	65.86	3.15	66.15	2.58

	ALL				
	Non-Handcrafted			Handcrafted	
	NET	ACC	SD	ACC	SD
RF	Inc. V3	72.48	1.21	68.41	1.40
K-NN	VGG	64.56	1.46	62.71	1.86
SVM	Inc. V3	71.49	1.30	51.73	0.77
MLP	Inc. V3	68.17	2.35	64.20	1.81

Experimental results

CLASSIFICATION RESULTS WITH FULLY-CONNECTED LAYERS

		InceptionV3		Incept. ResNetV2		ResNet50		VGG19	
		ACC	SD	ACC	SD	ACC	SD	ACC	SD
TASK 01	TL	67.74	9.40	63.34	8.64	51.61	4.56	66.45	10.35
	FT	69.68	5.86	67.74	7.90	70.32	11.50	**72.26**	**3.68**
TASK 02	TL	69.08	13.54	61.98	12.73	68.08	13.70	64.76	7.33
	FT	72.59	11.39	70.75	8.34	**75.17**	**8.63**	64.76	11.92

Discussion

- The results shows that non handcrafted features seem more promising than the handcrafted ones, reaching the best performance in terms of accuracy with the Random Forest classifier.
- The results obtained with handcrafted features are on the average worse than those obtained with non-handcrafted ones.
- For each task and for each classification scheme, there is always a CNN model whose features allow us to obtain better results than those obtainable with the handcrafted ones.
- Is noteworthy that handcrafted features contain also personal information, which are missing in the non-handcrafted ones, making the best non-handcrafted performances even more significant.
- The best result of the non-handcrafted feature is strengthened with the use of the fully connected layer with FT, which seems to produce the best of the best results.

Future works

- As future work we will consider the possibility to enrich the input image by using a ”multispectral” approach adding more channels associated to other dynamic information extracted from raw data. Furthermore, personal information (like age, education, etc.) may be added to the final set of
- non-handcrafted features.
- A second improvement should be obtained by including more than two handwriting tasks from the whole protocol.
- In this way should be possible to combine the classifier associated to each task with some kind of combination rule.