



# Feature Evaluation in Handwriting Analysis for Alzheimer's Disease Using Bayesian Network

Tiziana D'Alessandro, Claudio De Stefano , Francesco Fontanella ,  
Emanuele Nardone , and Alessandra Scotto di Freca <sup>(✉)</sup> 

Department of Electrical and Information Engineering, University of Cassino and Southern Lazio, 03043 Cassino, Italy

{tiziana.dalessandro,destefano,fontanella,emanuele.nardone}@unicas.it,  
a.scotto@uncias.it

**Abstract.** Alzheimer's disease, recognized as the most widespread neurodegenerative disorder worldwide, is intricately linked to cognitive impairments. The cognitive impairments, range from mild to severe and are a risk factor for Alzheimer's disease. They have profound implications for individuals, even as they maintain some level of daily functionality. In previous studies, it was proposed a protocol involving handwriting tasks as a potential diagnostic tool, and a comparative analysis of well-known and widely-used feature selection approach to determine the most effective features for predicting the symptoms related to cognitive impairments via handwriting analysis. In the presented study, we use a Bayesian Network to conduct further analysis of the most effective features extracted from handwriting. Our objective is to exploit the structural learning of Bayesian Networks to discover correlations among the most effective features for predicting impairment symptoms through handwriting analysis and deepen our understanding of the underlying cognitive functions affected. The results showed that the Bayesian Network chooses features conditionally dependent on the determination of the disease, and several features are selected more times than others, underlining their importance in the diagnosis. Moreover, comparing the results with those achieved by well-known and widely-used feature selection and classification approaches, the Bayesian networks exhibit the best performance by using a reduced set of features.

**Keywords:** Medical expert systems · cognitive impairments · Bayesian Networks · feature selection

## 1 Introduction

The World Health Organization recognizes that dementia is significantly underdiagnosed globally and emphasizes that even when a diagnosis is made, it often occurs at a relatively advanced stage. Delaying the onset of Alzheimer's disease

(AD) can have substantial benefits, including reduced care costs and increased lifespan for individuals. Therefore, early diagnosis of AD is crucial to enhance awareness and provide timely interventions. Mild cognitive impairment (MCI) is an early stage in the progression towards Alzheimer’s disease (AD); its accurate diagnosis is essential for initiating timely treatment to delay the onset of AD. Individuals experiencing MCI may start noticing alterations in their cognitive abilities, yet they are still capable of performing their daily tasks. However, severe levels of impairment can significantly impact the comprehension of events and the significance of information conveyed through speech and writing, ultimately leading to the loss of independent living. Currently, AD diagnosis relies on various methods such as imaging, blood tests, and lumbar punctures (spinal sampling), among others. Recent research has demonstrated that individuals with AD exhibit disrupted spatial organization and impaired motor control. Therefore, the assessment of motor activities, including the analysis of handwriting, which encompasses a complex interplay of cognitive, kinesthetic, and perceptual motor skills, holds significant potential in diagnosing AD. An illustrative example is the occurrence of dysgraphia in both the early and progressive stages of AD [8, 24]. Within this context, numerous studies in medicine and psychology have investigated the relationship between the disease and various handwriting features, utilizing conventional statistical methods [16, 19, 21]. However, these studies often neglect the intricate interactions that can arise among multiple features, failing to capture the complexity inherent in the analysis. In many instances, individual features that exhibit weak correlations with the target class could significantly enhance classification accuracy when combined with complementary features. Conversely, features that are individually relevant may become redundant when utilized alongside other features. In a previous study [4], the authors assessed the efficacy of the extracted features and their relationship with the diseases they potentially contribute to prediction. The techniques employed a search strategy to identify optimal solutions, i.e., the best feature subsets, based on a predefined evaluation function. The approaches used to define the best features are typically categorized into three main classes: filter, wrapper, and embedded methods. Filter methods primarily rely on statistical properties of the feature subset space. Wrapper methods, on the other hand, assess the performance of a specific classifier when utilizing a particular feature subset. Embedded methods incorporate feature selection as an integral part of the training process. In this previous work, the analysis of features extracted from the handwriting of individuals with neurodegenerative diseases and cognitive impairment is done using wrapper methods. In this paper, we present a novel approach based on Bayesian Networks to further investigate the complex interactions that may emerge among multiple features. A Bayesian Network (BN) is a probabilistic graphical model that encodes the joint probability of a set of variables that, in our case, can be the features and the disease identification. The feature selection problem has seen the extensive application of Bayesian networks [2, 18]. Once the BN has been learned from instances in a dataset, it allows the identification of a reduced set of features conditionally

dependent on the disease identification. Such a reduced set is also known as Markov Blanket (MB). Existing approaches involve learning a Bayesian network from the given dataset and subsequently utilizing the Markov Blanket of the target feature as the criterion for selecting relevant features. Thus, in this paper, we want to primarily explore the relationship between these diseases and individual features and then study the complex interactions that may emerge among multiple features. In this way, we can fill the gap in the literature regarding the comprehensive understanding of the interplay among various handwriting features in relation to AD and cognitive impairments (CI) [6]. The remainder of the paper is organized as follows. In Sect. 2 we describe the protocol we used to acquire the handwriting data, and the used features. Section 3 introduces the BNs and describes how they can be applied for selecting features. Finally, Sect. 4 reports the experimental results.

## 2 Acquisition Protocol and Features

We have developed a comprehensive protocol to collect data on handwriting movements from patients with CI and a control group of healthy individuals. The protocol, described in [4], consists of twenty-five tasks categorized as follows: graphic tasks, copy and reverse copy tasks, memory tasks, and dictation tasks. Graphic tasks assess the ability to write basic strokes, connect dots, and draw geometric shapes of varying complexity. Copy and reverse copy tasks evaluate the proficiency in reproducing complex gestures with semantic meaning, such as letters, words, and numbers. Memory tasks examine changes in the writing process for previously memorized words or words associated with depicted objects. Dictation tasks aim to investigate how handwriting performance is influenced by working memory usage.

It is important to note that each task was designed to assess either functional or parametric aspects. For example, in task number 17, participants were asked to write six different words that were analyzed in two different ways: the former, by averaging feature values across the entire word set, and the latter, by averaging feature values for each individual word. This led to the subdivision of the 17<sup>th</sup> task into six additional sub-tasks from 26<sup>th</sup> to 31<sup>th</sup>. A similar approach was applied to task number 14, which involved memorizing and rewriting the Italian words “telefono”, “cane”, and “negozio” added the sub-tasks 32, 33, and 34. The reasoning for this approach is to measure the impact of tiredness, i.e., if writing performance deteriorates more rapidly in participants suffering from neurodegenerative disorders when they are required to write multiple words consecutively. To summarize, we have a collection of thirty-four tasks that characterize each patient. The recruitment for the study uses standard clinical tests, including the Mini-Mental State Examination (MMSE) [11], Frontal Assessment Battery (FAB) [14], and Montreal Cognitive Assessment (MoCA) [17]. These tests assessed cognitive abilities across various domains, such as orientation, recall, and registration. To ensure unbiased results, the control group was carefully matched with the patient group regarding age, education level, gender, and

type of work (manual or intellectual), as shown in Table 1. Participants using psychotropic medication or any other drugs that could influence cognitive abilities were excluded. Moreover, we excluded patients with severely compromised cognitive abilities.

**Table 1.** Average demographic data of participants. Standard deviations are shown in parentheses.

	Age	Education	#Women	#Men
Patients	71.5(9.5)	10.8(5.1)	46	44
Control group	68.9(12)	12.9(4.4)	51	39

For data acquisition, we utilized a Wacom Bamboo Folio smart pad paired with a pen that allowed participants to write naturally on A4 white paper sheets placed on it. The smart pad recorded the x-y coordinates of pen movements (at a frequency of 200 Hz) on the paper’s surface. We also captured the pressure applied when the pen tip touched the sheet and the pen’s movements when lifted in the air within a maximum distance of 3 cm. The smart pad was positioned approximately seventy centimeters away from the participants during the data collection process. Importantly, all participants underwent the acquisition under identical conditions.

The features extracted from the raw data available, i.e., (x, y) coordinates, pressure, and timestamps, were calculated on the strokes making up the handwritten traits and then averaged over the entire task. Our goal is to describe, for each task, the behavior of a subject, taking into account a fixed number of features that are described in Table 2.

Considering the significant variation in the number of strokes across different subjects and tasks, we have adopted an averaging approach to consolidate the values extracted from individual strokes. Specifically, the feature denoted as  $f_{22}$  represents the total number of strokes. For each of the features from  $f_1$  to  $f_{21}$ , we have calculated the average and standard deviation, symbolized by  $f$  and  $\hat{f}$ , respectively. Consequently, the first 21 features are duplicated for each patient, encompassing static and dynamic handwriting characteristics. Additionally, features ranging from  $f_{23}$  to  $f_{26}$  consider variations associated with factors such as sex, age, work, and education.

As many studies in the literature show significant differences in patients’ motor performance between in-air and on-paper traits, each feature was calculated separately for the in-air or on-paper traits. In particular, we extracted four groups of features:

- **On-paper:** The features extracted from the written traits (i.e., during pen-down and the successive pen-up). Note that in this case, forty-seven features represented each sample.

**Table 2.** Feature list description.

$f_{ID}$	Name	Description
$f_1$	Duration	Time interval between the first and the last points in a stroke
$f_2$	Start Vertical Position	Vertical start position relative to the lower edge of the active digitizer area
$f_3$	Vertical Size	Difference between the highest and lowest y coordinates of the stroke
$f_4$	Peak vertical velocity	Maximum value of vertical velocity among the points of the stroke
$f_5$	Peak vertical acceleration	Maximum value of vertical acceleration among the points of the stroke
$f_6$	Start horizontal position	Horizontal start position relative to the lower edge of the active tablet area
$f_7$	Horizontal size	Difference between the highest (rightmost) and lowest (leftmost) X coordinates of the stroke
$f_8$	Straightness error	It is calculated by estimating the length of the straight line, fitting the straight line, estimating the (perpendicular) distances of each point to the fitted line, estimating the standard deviation of the distances, and dividing it by the length of the line between beginning and end
$f_9$	Slant	Direction from the beginning point to endpoint of the stroke, in radiant
$f_{10}$	Loop Surface	Area of the loop enclosed by the previous and the present stroke
$f_{11}$	Relative initial slant	Departure of the direction during the first 80 ms to the slant of the entire stroke
$f_{12}$	Relative time to peak vertical velocity	Ratio of the time duration at which the maximum peak velocity occurs (from the start time) to the total duration
$f_{13}$	Absolute size	Calculated from the vertical and horizontal sizes
$f_{14}$	Average absolute velocity	Average absolute velocity computed across all the samples of the stroke
$f_{15}$	Road length	Length of a stroke from beginning to end, dimensionless
$f_{16}$	Absolute y jerk	The root mean square (RMS) value of the absolute jerk along the vertical direction, across all points of the stroke
$f_{17}$	Normalized y jerk	Dimensionless as it is normalized for stroke duration and size
$f_{18}$	Absolute jerk	The Root Mean Square (RMS) value of the absolute jerk across all points of the stroke
$f_{19}$	Normalized jerk	Dimensionless as it is normalized for stroke duration and size
$f_{20}$	Number of peak acceleration points	Number of acceleration peaks both up-going and down-going in the stroke
$f_{21}$	Pen pressure	Average pen pressure computed over the points of the stroke
$f_{22}$	number of strokes	Total number of strokes of the task
$f_{23}$	Sex	Subject’s gender
$f_{24}$	Age	Subject’s age
$f_{25}$	Work	Type of work of the subject (intellectual or manual)
$f_{26}$	Education	Subject’s education level expressed in years

- **In-air:** The features extracted from the in-air traits. These movements characterize the planning activity for positioning the pen tip between two successive written traits. Note that in this case, we extracted forty-five features because pressure (feature  $f_{21}$ ) is always zero.
- **All:** In this scenario, each feature vector includes in-air ( $af_i$ ) and on-paper ( $pf_i$ ) attributes (where the subscript  $i$  indicates the feature number), thus reporting the values of both On-paper and In-air feature vectors. The aim was twofold: facilitating a direct in-air versus on-paper feature comparison and delving into the intricate interplay between the two. It is worth noting that eighty-eight distinct features were considered, excluding repeated personal features and pressure variables regarding the in-air part.
- **In-air-On-paper:** The computation of these features disregards the differentiation between in-air and on-paper characteristics. In practice, the value of each feature is obtained by averaging the values derived from both in-air

and on-paper traits. The only exception is for the pressure, whose values are obviously obtained considering on-paper traits. This approach represents an alternative method of supplying global information to the classification system, considering as equivalent the motor planning for both handwritten and in-air strokes.

Summarizing, we considered four categories of features: we have forty-five features for In-air category, forty-seven features for both On-paper and In-air-On-paper categories, and eighty-eight for the category All.

### 3 Bayesian Network for Feature Evaluation

The problem of feature evaluation can be handled by estimating the joint probability of each feature and the class label. A Bayesian Network (BN) may effectively solve this problem. A BN is a probabilistic graphical model that allows the representation of a joint probability distribution of a set of random variables through a Direct Acyclic Graph (DAG) [20]. The graph nodes represent the variables, while the arcs encode the statistical dependencies among the variables. An arrow from a node  $f_i$  to a node  $f_j$  encodes the conditional dependence between the node  $f_j$  and node  $f_i$ , and we can define  $f_i$  as a *parent* of  $f_j$ . In a BN, the  $i$ -th node  $f_i$  is associated with a conditional probability function  $p(f_i|pa_{f_i})$ , where  $pa_{f_i}$  indicates the set of nodes which are parents of  $f_i$ . Such a function quantifies the effect that the parents have on that node. The process of learning a BN entails acquiring knowledge from a training set of examples. This learning phase involves capturing both the network structure, which defines the statistical dependencies among variables, and the parameters of the probability distributions associated with those variables. Among structural learning methods, there are constraint-based methods like PC [22], IAMB [23] that exploit conditional independence relationships in the data to uncover the network structure; there are also score-based methods that evaluate different network structures based on a scoring metric to find the structure with the highest score or the lowest complexity. Among score-based methods, there are K2 [5], TAN [12] etc. The third category of structural learning methods, the hybrid ones combine the strengths of constraint-based and score-based approaches. These methods balance computational efficiency and the ability to handle complex network structures. Among them, there are methods also based on evolutionary algorithm [7]. On the other hand, parameter learning generally uses the Maximum Likelihood Estimation that estimates the parameters of a BN by maximizing the likelihood of the observed data. Once the statistical dependencies among variables have been learned, the DAG structure encodes them, and the joint probability of the set of variables  $F = \{f_1, \dots, f_L\}$  can be described as:

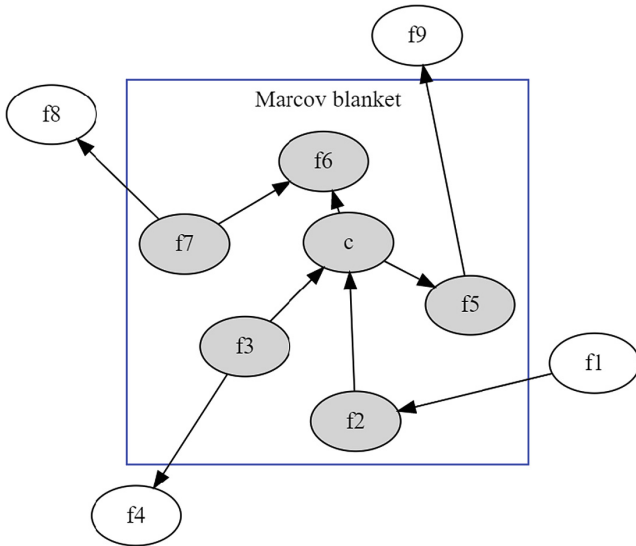
$$p(f_1 \dots, f_L) = \prod_{f_i \in F} p(f_i|pa_{f_i}) \quad (1)$$

In the feature evaluation framework, this property can be used to infer the true class  $c$  of an unknown sample only by a subset of features. In fact, suppose

to have  $L$  features, then the class label  $c$  and the  $L$  features can be modeled as a set of  $(L + 1)$  variables  $\{c, f_1, \dots, f_L\}$ , and the Eq. (1) allows the description of their joint probability as:

$$p(c, f_1, \dots, f_L) = p(c | pa_c) \prod_{f_i \in F} p(f_i | pa_{f_i}) \tag{2}$$

The node  $c$  may be the parent of one or more of the nodes of the DAG. For example, if we consider the BN depicted in Fig. 1, we have that  $c$  is the parent of nodes  $f_6$  and  $f_5$ . While the nodes  $f_2$  and  $f_3$  are the parents of  $c$ . Therefore, it may be useful to divide the set of DAG nodes that are not parents of  $c$  into two groups: the first, denoted  $F_c$ , contains the nodes having the node  $c$  among their parents, and the second, denoted  $F_{\bar{c}}$ , the remaining ones. Note that among  $F_c$  nodes there also are nodes, like,  $f_7$  that are not parents of  $c$ , but are modeled in the conditional probability that also contains the node  $c$ . With this assumption, Eq. (2) can be rewritten as:



**Fig. 1.** An example of Bayesian Network and the Markov Blanket in case of 9 features available.

$$p(c, f_1, \dots, f_L) = p(c | pa_c) \prod_{f_i \in F_c} p(f_i | pa_{f_i}) \prod_{f_i \in F_{\bar{c}}} p(f_i | pa_{f_i}) \tag{3}$$

This property allows a BN to recognize a given sample only considering the responses provided by the feature represented by the nodes that are directly linked to the class node. The group  $F_c$  and  $pa_c$  is also known as Markov Blanket (MB) of node  $c$  and in  $F_{\bar{c}}$  group are all nodes d-separated from  $c$ , that is, it

contains features conditionally independent from class label  $c$ . Therefore, the MB of node  $c$  consists of its parents, children, and spouses, and the  $c$  is independent of all other nodes given its MB.

This behavior encoded by the BN is particularly useful in the testing phase when, from the feature vector of all features, the class label is inferred. In fact, given the set of responses concerning a sample, we can use the conditional probability  $p(c|f_1, \dots, f_L)$  estimated by the BN for assigning the most probable class  $\hat{c}$  to the unknown sample, as follows:

$$\hat{c} = \arg \max_{c \in C} p(c|f_1, \dots, f_L) \quad (4)$$

where  $C$  is the set of classes. Considering the definition of conditional probability, and omitting the terms not depending on the variable  $c$ , the above equation can be rewritten as follows:

$$\hat{c} = \arg \max_{c \in C} \frac{p(c, f_1, \dots, f_L)}{p(f_1, \dots, f_L)} = \arg \max_{c \in C} p(c, f_1, \dots, f_L) \quad (5)$$

which involves only the joint probabilities  $p(c, f_1, \dots, f_L)$ . According to Eq. (3), and discarding the term conditionally independent on  $c$ , Eq. (5) assumes the form:

$$\hat{c} = \arg \max_{c \in C} p(c|pa_c) \prod_{f_i \in F_c} p(f_i|pa_{f_i}) \quad (6)$$

An example of the application of this rule is shown in Fig. 1, where only 9 features have been considered for the sake of clearness. In this case, the above equation becomes  $\hat{c} = \arg \max_{c \in C} p(f_6|c, f_7)p(f_5|c)p(c|f_3, f_2)$ . Then in the case of the learned BN of Fig. 1 we have to know only features  $f_2, f_3, f_5, f_6$  and  $f_7$  to infer  $c$ . In fact, during the learning procedure, the set of experts  $F_{\bar{c}}$ , which does not add information to the choice of  $\hat{c}$ , is individuated and it is discarded in the testing phase. In our example, in fact, the contribution of features  $f_1, f_4, f_8$  are considered not necessary, and they can be discarded in the testing phase. Thus, the BN-based feature selection approach uses only the features in the MB of node  $c$ .

## 4 Experimental Findings

In this section, we will describe the experimental setup and procedures implemented to assess the performance of our system. The data were acquired according to the protocol described in Sect. 2 and refer to 18 subjects, including 90 patients and 90 healthy people. The purpose of the experimentation was twofold: to show that our approach allows us to accurately classify patients and healthy subject using the features extracted fro their handwriting, and to underline which features are more relevant to provide the diagnosis. In order to achieve our aim, we learned the BN DAG structure and the conditional probabilities among features and the class label, using the available data. Given the DAG structure,

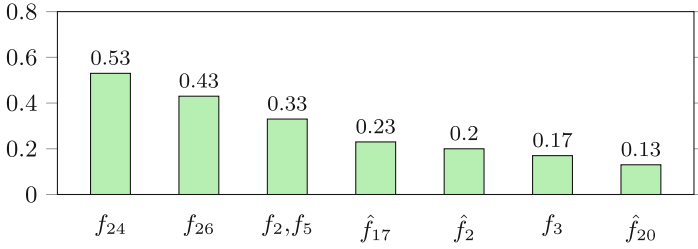


we extracted the MB, which highlights the features conditionally dependent on the class node. We performed the aforementioned procedure by running the K2 algorithm 30 times, setting to 3 the maximum number of parents for each node. The value selected for the maximum number of parents represent a compromise between algorithm efficiency and computational cost, while the number of runs of the algorithm k2 was chosen to average the effects of the initial ordering of the input variables. Its efficiency, in fact, is strongly dependent on this ordering and therefore a different initial order of the variables were considered in each run. We used the WEKA software for feature selection and classification, whereas we used the KNIME software for data pre-processing (managing missing values, filtering null columns, and encoding categorical variables). Although real-world data often includes a combination of discrete and continuous variables, BN structure learning algorithms generally assume that all random variables are discrete. As a consequence, continuous variables are typically discretized to comply with this assumption. In our implementation, we applied the sample quantile technique to discretize continuous variables into five binning intervals. These pre-processing steps were crucial in preparing the data for the subsequent classification analysis.

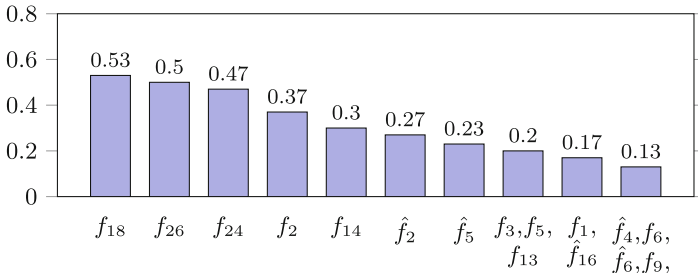
In order to investigate the importance of the used features, we plotted the histogram of the features selected more than 10% of the times in the MB among the 30 runs (see Fig. 2) for each category, namely In-air, On-paper, In-air-On-paper and All. Even if the results relative to the feature category All indicate that the features derived from on-paper traits are selected more frequently than those related to on air traits, the importance of on air traits is confirmed by our results reported in Table 3 as well as by the results reported in [9]. From the figure, we can see that the age, feature  $f_{24}$  (see Table 2), is the most selected in the four categories, with a minimum value of 0.47 (On-Paper). This confirms that age significantly affects the handwriting process of people with Alzheimer’s disease, due to the changes in brain structure, such as the decrease in the size of the brain’s memory center (hippocampus). Even if these changes typically worsen with age [10], they are more relevant in subjects with mild cognitive impairment and even more dramatic in people with Alzheimer’s disease. Another feature selected with high frequency in 3 of the 4 categories is the education, feature  $f_{26}$ , with a minimum value of 0.43. Comparing these results with the ones obtained in [4], where the education was selected few times, we can say that BNs seem more effective in estimating the correlation among variables, and thus the joint contribution of groups of features in distinguishing patients from healthy people. Other two significant features selected with high frequency are  $f_2$  and  $f_3$ , measuring the start vertical position and the vertical size, respectively. The joint selection of these features confirms the importance of evaluating the spatial coordination of subjects, and indicates that their variability differs between healthy people and AD patients. As regards the evaluation of the dynamic parameters of the handwriting process, separate evaluations can be made for data On-Paper and In-Air data. The peak vertical acceleration  $f_5$  appears with its mean value and its mean and standard deviation in On-Paper group, underlining the importance of the variability of this feature. In In-Air group, the presence of  $f_5$  together

with the standard deviation of the number of peak acceleration points  $\hat{f}_{20}$  and the standard deviation of the normalized y jerk  $\hat{f}_{17}$ , suggests that the variability of in-air movements may highlight anomalies in the handwriting of AD patients. On the other hand, also the movements performed with the pen tip touching the paper are characterized by different dynamics in AD patients and healthy people. In fact, in the On-Paper category, the mean value of jerk  $f_{18}$ , the absolute y jerk  $f_{16}$ , the average absolute velocity  $f_{14}$  and the standard deviation of peak vertical acceleration  $\hat{f}_4$ , are the most selected features, underlining the importance of hesitations in the handwriting process. In the On-Paper category, features like  $f_3$ ,  $f_{13}$ ,  $f_1$ ,  $f_6$ , and  $f_9$  that measure the space occupation, are also very important. In case of In-Air-On-Paper group, the features are obtained by averaging the values from both in-air and on-paper attributes for each task, thus assuming that the generation of the handwritten strokes is obtained by concatenating in air and on paper movements. Apart from the already mentioned features, the most important features are the absolute jerk  $f_{18}$ , peak vertical velocity  $f_4$ , the road length  $f_{15}$ , but also the duration  $f_1$ , the average absolute velocity  $f_{14}$ , the start horizontal  $f_6$  and vertical position,  $f_2$ , with their mean and standard deviation. These results confirm that globally handwriting dynamic and spatial coordination are very important. When we apply the BNs to the category All, the effect is to produce a ranking of the features computed on both In-Air and On-Paper traits. In particular, we have the predominance of On-Paper features, where the most selected features are pen pressure  $f_{21}$ , the pen absolute velocity  $f_{14}$ , the absolute y jerk  $f_{16}$ , the start vertical position  $f_2$ , loop surface  $f_{10}$ , the road length  $f_{15}$ , duration  $f_1$ . The only feature computed on in-air traits present in the histogram is the feature the vertical size  $f_3$ . It is interesting to note that the features  $f_{10}$  emerge only in this group, meaning that they probably assume importance only in correlation with other features of both In-Air and On-Paper category. Finally, note that the feature road length  $f_{15}$  is present in all the histograms except for the one relating to the in-air category.

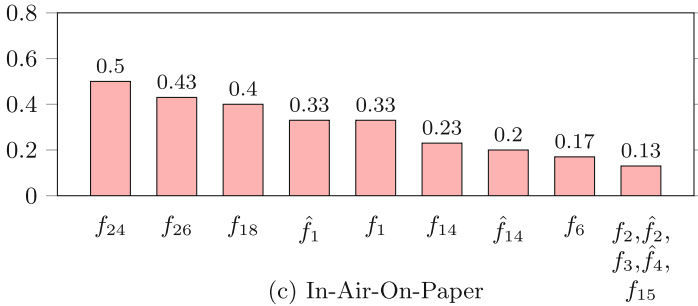
We used the Recursive Feature Elimination with Cross-Validation (RFE) [13] to select the most relevant features for comparison purposes. This technique recursively eliminates features and evaluates their impact on the performance of basic classifiers. We used a 10-fold cross-validation setup and evaluated the performance using the XGBoost [3], Decision Tree [15], and Random Forest [1] classifiers. We performed 30 classification runs, each time using the relevant features selected by each method. Table 3 shows the results obtained in all the aforementioned experiments; the first column shows the algorithm used for the feature selection, whereas the second column shows the algorithm used for the classification step. For each method, we reported the accuracy mean and the standard deviation along the 30 runs, with the average mean number of features selected (NF). From the table, we can observe that the best results are obtained every time with BN selection and classification. It is also worth noticing that when we apply another classification method to the feature selected by BNs it never achieved the result obtained by using BN as a classifier. This behavior proves that both structural and parametric learning of the BN is very effective



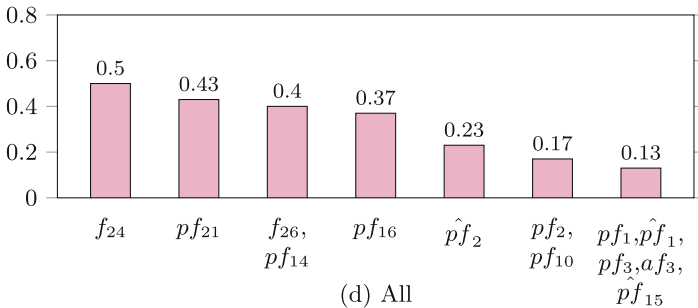
(a) In-Air



(b) On-Paper



(c) In-Air-On-Paper



(d) All

**Fig. 2.** Feature percentage greater than 10% selected by the proposed approach. Each bar of the histogram shows the fraction of time the corresponding feature(s) was selected among the 30 runs.

in selecting features and classifying healthy subjects and AD patients. Moreover, among the four categories, the best performing is All and then, in the order, In-air-On-paper, On-paper and In-air, and This result shows that BN optimally exploits the information coming from the two groups when this information is not averaged. Furthermore, as also shown in Fig. 2(d), the best features are those derived by the On-Paper category, where the In-Air ones are exploited as complementary information to obtain the best recognition performance.

**Table 3.** Classification results using BN, RF, XGBoost, and DT in term of average accuracy (Acc) and its standard deviation (in brackets), and average number of selected features (NF).

Feature Selection	Classifier	All		In-Air		In-air-On-paper		On-Paper	
		Acc	NF	Acc	NF	Acc	NF	Acc	NF
BN	BN	74.99 (5.74)	48.13	71.22 (4.03)	25.27	73.26 (4.47)	24.77	73.11 (4.11)	23.73
	RF	61.34 (4.76)	48.13	61.53 (4.72)	25.27	61.15 (4.24)	24.77	59.57 (3.68)	23.73
	XGB	57.20 (4.88)	48.13	58.73 (5.64)	25.27	58.69 (4.45)	24.77	56.91 (5.31)	23.73
	DT	64.31 (6.66)	48.13	58.12 (6.53)	25.27	63.52 (7.24)	24.77	63.32 (6.89)	23.73
RFE	RF	61.90 (4.26)	17.03	62.15 (3.93)	19.09	61.93 (4.81)	22.41	61.56 (5.36)	14.38
	XGB	61.02 (5.69)	11.90	60.28 (4.47)	10.60	59.94 (6.16)	11.20	59.86 (5.41)	10.40
	DT	63.39 (6.38)	10.25	62.56 (5.39)	11.38	64.63 (4.44)	10.25	64.28 (4.69)	10.19

## 5 Conclusions

In this study, we presented a novel approach based on Bayesian networks to evaluate the statistical dependencies among different features extracted from handwriting samples, in order to maximize the performance of a system for early AD diagnosis. The data were obtained by administering handwriting tests according to a protocol including 35 tasks, to a group of 180 subjects including 90 healthy controls and 90 AD patients. From these data, four datasets were obtained including feature relative to on paper and on air traits: this choice allowed us to estimate the distinctive power of the different considered feature categories and to study the complex interactions among groups of features.

The results seem very encouraging and demonstrate the effectiveness of the proposed approach: in particular, the Bayesian network allowed the selection of about half of the whole set of available features, significantly improving the performance with respect to other state of the art feature selection methods. As future works, we plan to increase the number of parents in the BN structural learning algorithm and to evaluate the sensitivity of the proposed system to variations of this parameter. The number of parents, in fact, has a very strong impact on the computational cost of the BN learning algorithms. We also plan to apply hybrid structural learning techniques to Bayesian networks [7].

**Acknowledgements.** The research leading to these results has received funding from Project “Ecosistema dell’innovazione - Rome Technopole” financed by EU in NextGenerationEU plan through MUR Decree n. 1051 23.06.2022 - CUP H33C22000420001.

## References

1. Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001)
2. Castro, P.A.D., Von Zuben, F.J.: Learning Bayesian networks to perform feature selection. In: 2009 International Joint Conference on Neural Networks, pp. 467–473 (2009)
3. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2016, pp. 785–794. ACM, New York (2016)
4. Cilia, N.D., De Stefano, C., Fontanella, F., Scotto di Freca, A.: Feature selection as a tool to support the diagnosis of cognitive impairments through handwriting analysis. *IEEE Access* **9**, 78226–78240 (2021)
5. Cooper, G.F., Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data. *Mach. Learn.* **9**(4), 309–347 (1992)
6. De Stefano, C., Fontanella, F., Impedovo, D., Pirlo, G., Scotto di Freca, A.: Handwriting analysis to support neurodegenerative diseases diagnosis: a review. *Pattern Recognit. Lett.* **121**, 37–45 (2019)
7. De Stefano, C., Fontanella, F., Marrocco, C., di Freca, A.S.: A hybrid evolutionary algorithm for Bayesian networks learning: an application to classifier combination. In: Di Chio, C., et al. (eds.) *EvoApplications 2010*. LNCS, vol. 6024, pp. 221–230. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-12239-2\\_23](https://doi.org/10.1007/978-3-642-12239-2_23)
8. Di Mauro, G., Bevilacqua, V., Colizzi, L., Di Pierro, D.: Testgraphia, a software system for the early diagnosis of dysgraphia. *IEEE Access* **8**, 19564–19575 (2020)
9. Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z., Faundez-Zanuy, M.: A new modality for quantitative evaluation of parkinson's disease: in-air movement. In: 13th IEEE International Conference on BioInformatics and BioEngineering, pp. 1–4 (2013). <https://doi.org/10.1109/BIBE.2013.6701692>
10. El-Yacoubi, M.A., Garcia-Salicetti, S., Kahindo, C., Rigaud, A.S., Cristancho-Lacroix, V.: From aging to early-stage Alzheimer's: uncovering handwriting multimodal behaviors by semi-supervised learning and sequential representation learning. *Pattern Recognit.* **86**, 112–133 (2019). <https://doi.org/10.1016/j.patcog.2018.07.029>. <https://www.sciencedirect.com/science/article/pii/S0031320318302693>
11. Folstein, M.F., Folstein, S.E., McHugh, P.R.: 'Mini-mental state': a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatric Res.* **12**(3), 189–198 (1975)
12. Friedman, N., Geiger, D., Goldszmidt, M.: Bayesian network classifiers. *Mach. Learn.* **29**(2–3), 131–163 (1997)
13. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. *Mach. Learn.* **46**(1–3), 389–422 (2002)
14. Iavarone, A., Ronga, B., Pellegrino, L., Loré, E., Vitaliano, S., Galeone, F., Carlomagno, S.: The frontal assessment battery (FAB): normative data from an Italian sample and performances of patients with Alzheimer's disease and frontotemporal dementia. *Funct. Neurol.* **19**, 191–195 (2004)
15. L. Breiman, J. Friedman, R.O., Stone, C.: Classification and regression trees. Wadsworth (1984)
16. Müller, S., Preische, O., Heymann, P., Elbing, U., Laske, C.: Diagnostic value of a tablet-based drawing task for discrimination of patients in the early course of Alzheimer's disease from healthy individuals. *J. Alzheimers Dis.* **55**(4), 1463–1469 (2017)

17. Nasreddine, Z.S., et al.: The Montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* **53**(4), 695–699 (2005)
18. Njah, H., Jamoussi, S., Mahdi, W.: Interpretable Bayesian network abstraction for dimension reduction. *Neural Comput. Appl.* **35**(14), 10031–10049 (2023)
19. Onofri, E., Mercuri, M., Salesi, M., Ricciardi, M., Archer, T.: Dysgraphia in relation to cognitive performance in patients with Alzheimer’s disease. *J. Intellect. Disabil. Diagn. Treat.* **1**, 113–124 (2013)
20. Pearl, J.: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, Burlington (1988)
21. Small, J., Sandhu, N.: Episodic and semantic memory influences on picture naming in Alzheimer’s disease. *Brain Lang.* **104**(1), 1–9 (2008)
22. Spirtes, P., Glymour, C., Scheines, R.: *Causation, Prediction, and Search*. MIT Press, Cambridge (2000)
23. Tsamardinos, I., Aliferis, C.F., Statnikov, A.R., Statnikov, E.: Algorithms for large scale Markov blanket discovery. In: *FLAIRS Conference*, St. Augustine, FL, vol. 2, pp. 376–380 (2003)
24. Yan, J.H., Rountree, S., Massman, P., Doody, R.S., Li, H.: Alzheimer’s disease and mild cognitive impairment deteriorate fine movement control. *J. Psychiatr. Res.* **42**(14), 1203–1212 (2008)