POLITECNICO
MILANO 1863

DEPARTMENT OF ELECTRONICS , INFORMATION
AND BIOENGINEERING
DOCTORAL PROGRAMME IN BIOENGINEERING

Novel eHealth methodologies to support the detection of age related decline

Advisor
Prof. **Simona Ferrante**
Co-advisers
Francesca Lunardini phd
Tutor
Prof. **Raffaele Dellaca'**
Chair of the Doctoral Program:
Prof. **Andrea Aliverti**

DOCTORAL DISSERTATION by
Davide Di Febbo

2021-XXXIII Cycle

Abstract

Nowadays, and more in the next future, the ageing population is determining a large impact on the public healthcare systems worldwide. To face the growing demand for long term services, the current healthcare model needs to shift its focus on the prevention and the early identification of chronic diseases, whose incidence is high in people aged 65+. Timely interventions are key to slow down decline and mitigate the symptoms in older individuals, besides limiting the use of hospital resources. Home monitoring technologies can allow the remote examination of the patients and the continuous tracking of the seniors' health status in their living environment. Therefore, early signs of decline could be recognised more quickly. However, some issues related to the acceptance of the daily monitoring systems by the users and to the data reliability in uncontrolled setting could represent a barrier for their effective application. The main objective of my PhD is to design and develop novel eHealth solutions such as IoT sensors, explainable artificial intelligence applications and decision support systems to enable the early detection and remote assessment of decline in older adults. In particular two solutions are here proposed: the exploitation of the ecological assessment of handwriting as daily life activity monitoring and a decision support system for the posterior interpretation and the evaluation of a complex neuro-psychological test. These technologies were successfully tested and validated using specific protocols and data which simulated the real-application scenario. The remarkable results suggested that these may be promising solutions for detecting physical and cognitive decline in the home setting.

Summary

Introduction

One of the today's major challenges for the healthcare systems worldwide is represented by the ageing population. Although it is a great achievement for our society, the shift towards the older age of the population will require a huge demand of resources and long term care [1]. From the age of 60, the incidence of various chronic diseases, such as frailty, Parkinson's disease and dementia, rapidly increases and, if not timely treated, they can seriously impact the individual's quality of life. To date, early diagnosis represents an effective strategy to slow down and mitigate the physical or cognitive degradation process lead by a chronic disease [1, 2]. However, ageing is a complex and multi-factorial process and the trajectories of decline can consistently differ between individuals. In their early stages, the phenotype of some age-associated pathological conditions can be confused with the typical signs of healthy ageing [3]. Moreover, in some conditions as frailty, there is a lack of standard tools in medicine for an effective early diagnosis [4]. Chronic diseases in older adults generally start in mid-age with a pre-clinical stage, in which the individual experiences a slight physical or cognitive decline [5]. The symptoms which could be associated to a certain disease begin to emerge in the mild clinical stage of decline, even after 20 years from the pre-clinical phase onset. Then, the course of decline progresses to the more severe clinical stage, in which the cognitive and/or physical functionality of the patient results markedly compromised. In most cases, only interventions in the pre-clinical stage can stop or significantly decrease the evolution of decline in the individual.

The health care provider's strategy should be focused to the early detection of decline in older adults, so that to identify the presence of a precursory stage of the pathology and provide the appropriate intervention. However, the current healthcare model is not well suited to this purpose. Historically, healthcare systems are optimised to tackle with acute conditions, by recognising the symptoms and managing the best interventions to the patient [1]. The health status of a patient is assessed through in-person spot visits, in which the symptoms are examined. Yet, a limited number of consultations in

time is not effective to identify the early signs of a potential abnormal decline in the subject. In addition, the non healthy decline in seniors is typically due to co-morbidity and symptoms might be non-communicable [6].

The increase of the life's expectancy in the present days is calling for new requirements for the healthcare systems, more focused on the prevention, the early detection and the efficient management of chronic diseases. The main effort should be dedicated to the development of more agile strategies for the assessment of the subject's health status. For example, the continuous monitoring of the individual's decline trajectory and remote consultations could help the healthcare systems to provide the right interventions over time, without the need of in-person visits. Remote health may also allows ageing in place, or rather possibility to live in one's own home and community safely, independently, and comfortably, regardless of age, income, or ability. In the home monitoring scenario, some clinical tests could be remotely administered for an agile assessment of the subject's particular health conditions. To this purpose, supervised digital tests and clinical expert systems can be useful to manage and evaluate the tests with more efficacy [7].

Technology plays the major role in the transformation of the healthcare system in a more integrated and person centred institution. In particular, telemedicine and internet of things (IoT) offer important instruments for the remote monitoring of older patients in home environments. Both the older individuals and long term-care systems benefit from the possibility to remotely identify degeneration in the domestic environment. A remote and continuous monitoring would be more effective than spot visits in detecting early signs of physical or cognitive decline and it would not require the same time and resources.

Researchers studied the use of smart home technology, mobile and wearable devices for the domestic monitoring of the community dwelling elders. Environmental sensing, for example presence, weight and temperature senors, can be used to rapidly locate the older individual inside the house and intervene in the case of domestic accidents or fall. Wearable systems may be used to track biometric quantities and mobile devices, such as smartphones and tablet, could be used to remotely access health services, contact the clinicians and execute routine checks. Recent studies investigated the validity of self-administered digital cognitive tests, implemented on mobile devices, for an agile assessment of the mental decline in seniors living alone. The outcome of those tests can be forwarded to the clinicians to help the early diagnosis of mental disorders, without the need for the elderly to often reach the hospital.

In the field of remote monitoring, research is still underway. Standard metrics need to be generally accepted and widely adopted to be used consistently. Moreover, in the case of older adults, monitoring technologies should issues related to the diffuse senior's technological illiteracy. The goal

is therefore to develop non-intrusive (or ecological) solutions which do not request elderly users to adopt new behaviours or perform tasks that may results uncomfortable. The instrumental daily-life activity (IADL) monitoring is an alternative that can fit these requirements, since it may retrieve health information about the user by carrying out a daily activity. However, IADL is operated in a generally unconstrained setting that might influence and deviate the measurement from the standard reference, if any. Information about decline from the daily-activities monitoring is therefore to be found in abnormalities in the data (for example, outliers and changes in the measurements).

Below, the general aim and the main objectives of this work are listed. In the following sections, the topics are treated more in the details.

This thesis addresses the problem of the detection of age-related decline.

The core of this research is dedicated to the development of the IADL monitoring through the ecological assessment of handwriting. With a transnational approach, the problem is addressed by transferring the main scientific findings in literature into a real-world scenario application, by developing novel ecological technologies and methodologies.

The other topic concerns the remote the remote analysis of a cognitive test. In this research, a decision support system for for the semi-automatic analysis of the Rey-Osterrieth complex figure test is designed and tested on retrospective data.

Ecological Monitoring of Handwriting

Handwriting has been largely investigated in literature and it has been proved to be a potential biomarker for decline in older adults and for various age-related pathological conditions [8]. The useful information is retrieved using quantitative analysis of handwriting, which consists in the calculation of specific measurements from the written trace kinematics, gesture dynamics and tremor that show significant variations with age and pathology [9, 10]. In most of the studies, these quantities (i.e. handwriting indicators) were measured using special tools, typically found on the market. Devices, such as tablets and digitizing surfaces have been used to collect handwriting data, following precise experimental protocols.

Being a common daily-activity, handwriting is well suited to be remotely assessed in the home environment. However, the translation of the previous findings in research to the home monitoring is not straightforward. The first barrier is represented by the instrumentation which can be used to acquire

the handwriting data. Tablets and digitizers might result user-friendly for younger adults, but they are certainly less so for the seniors which prefers the traditional ink-and-paper writing. The second issue lies in having to avoid, as far as possible, the data acquisition protocols to make monitoring more ecological. This setting represents an uncontrolled environment, unprecedented in literature, in which measurements are more likely to be noisy and the age or pathological reference values may not be valid anymore [11]. The specific objectives for this topic are:

i) provide a novel instrument to allow the ecological assessment of handwriting, ii) test the validity and the reliability of the handwriting indicators obtained from the data collected using that instrument and iii) design an anomaly detection approach to identify age and pathological variations in older adults' handwriting, for monitoring purpose.

The Smart Ink pen

This research begins with the design and development of a novel device for the ecological handwriting assessment. This activity was completed within the European project MoveCare (H2020, GA no. 732158) [12]. My specific contribution regarded the definition of the technical specifications (hardware and software), the testing and the validation of the device.

Since it was mainly meant to be used by seniors, ease of use was the first concern. The solution consisted of an instrumented ink pen, equipped with an inertial measurement unit (IMU) and a force sensor on the tip, which allowed the data collection during common paper-and-pen writing tasks. Figure 1 shows a computerised representation of the pen and its components. The sensors, the electronics, the rechargeable battery and all the

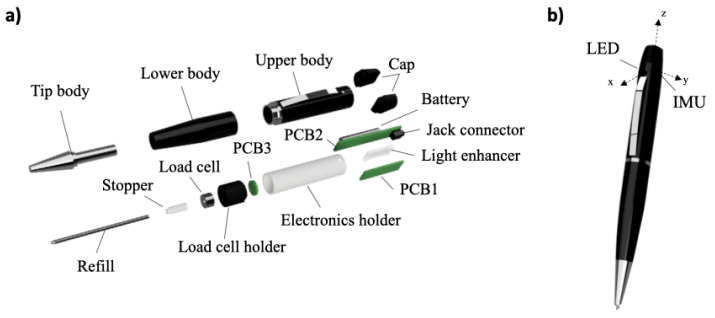

Figure 1: (a) A rendering image of the smart ink pen and its internal components. (b) An external view of the smart ink pen.

internal components were miniaturised and hidden by the pen body, not

very different from a common stylus. Similarly, the firmware was designed to autonomously start the data collection when the instrument was used to write and then send them via low-energy Bluetooth (BLE). A remote-controlled data recording modality was also included. The signals collected during handwriting were 3-axis linear accelerations, 3-axis angular velocities and the normal force applied to the pen tip.

Once designed and realised the device prototypes, my role consisted in devising and performing the testing and validation procedures. A series of tests were conducted to test and validate the signals acquired using the smart pen. At first, the static tip force was calibrated using and external force sensor. The force measurements, from the pen sensor and from the external load cell, were obtained applying testing weights (from 10 to 500g) above the stylus. The linear regression between the two sets of measurements returned a R^2 score of 0.99 and the linear coefficient were used to for the conversion of the non-scaled units of the pen sensor to grams. Then, the writing force was validated in dynamic conditions, i.e. during handwriting. The pen force and the external reference signals were synchronised and compared using the Pearson's correlation coefficient, with a mean value of 0.96.

The dynamic inclination of the pen was computed from the inertial measures. The tilt angle was validated using an optoelectronic inclination measurement of the pen attitude as reference. The comparison of the signal showed positive results with a correlation coefficient ranging from 0.89 to 0.78 and a root mean squared error (RMSE) between 3.8° and 6.3°.

Before the calculation of the handwriting indicators was made, the segmentation in strokes (i.e. the identification of the writing interval in the signals) was evaluated. The non-zero force tracts in the pen and the external load cell force signals were identified and compared during handwriting. The linear regression between the strokes duration reported an R^2 score of 0.99 and the agreement of the measurements was checked through the Bland-Altman analysis. Figure 2 shows the comparison between the writing force signal acquired with the pen and the reference signal from the external sensor in the upper panel. Below the signals, the writing tracts are highlighted by line segments. In the lower panel, the Bland-Altman plot for the writing tracts duration is reported.

In the Movecare project, a eHealth platform for the monitoring and assistance of the community dwelling elders was developed. The study was finalised to an ecological momentary assessment in which the platform was tested in the real-world scenario with 25 older adults (65 or more years old) living alone. The pilot study was organised in two rounds (round 1 and 2), each of them lasting 3 months and usability and acceptance of the various components of the eHealth platform were investigated in two main ways: i) through users' feedback collected using questionnaires; ii) by measuring

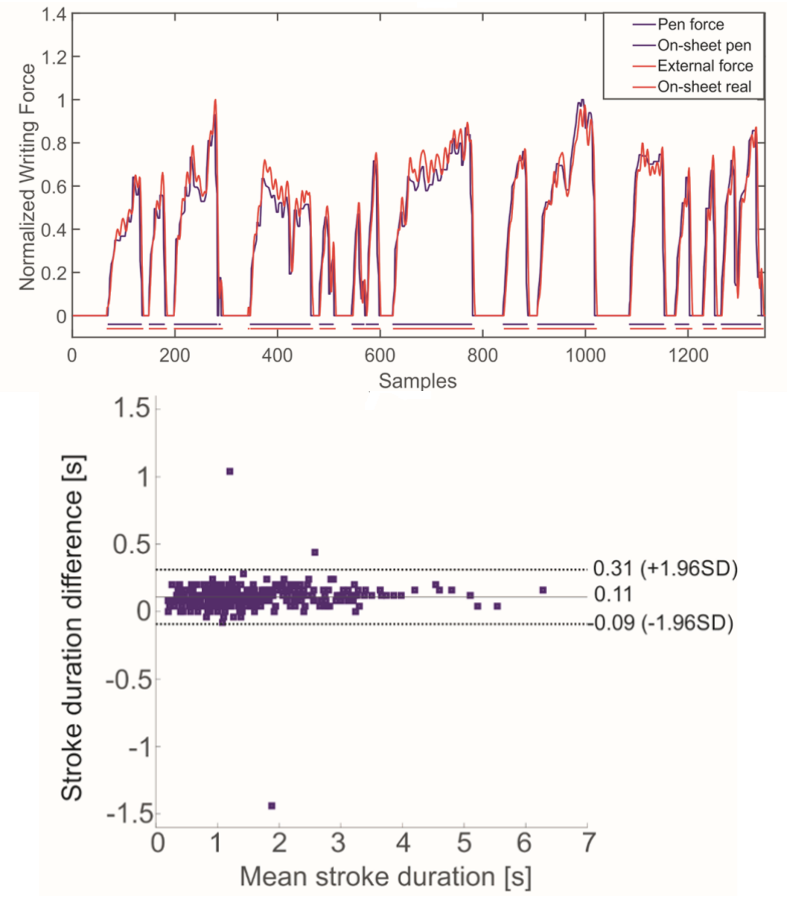

Figure 2: Pen vs external sensor's writing force signals and strokes duration (upper panel). Bland-Altman plot for the strokes duration (lower panel).

the amount of time the devices/modules were used. From the first set of questionnaires, the pen resulted well accepted by the recruited users. The usability scores obtained by the pen were very high: 90% of the participants rated the question *using the smart pen was easy* positively and the 86.4% rated the question *I found it easy to charge the smart pen* positively. The pen in particular resulted indeed the most used component of the Movecare monitoring platform.

Handwriting Indicators Reliability and Age-Related Differences

The handwriting assessment consists in the calculation and the analysis of some quantitative indicators which describe particular characteristics of the writing trace kinematics, the gesture dynamics and the hand tremor [9, 13]. In literature, the indicators have been observed to be sensible to variations in handwriting due to a normal or pathological ageing decline process [8, 14, 15, 10]. In these researches, handwriting have been studied in controlled conditions, i.e. using standard writing protocols or drawings which have been largely tested. Indeed, the home environment represents an new uncontrolled setting to acquire handwriting data. Therefore, with the aim of extending the ecological validity of the handwriting assessment to the remote setting [11], a special protocol which resembled the common daily writing activity have been designed: the subject were asked to write a short free text and a grocery list without any constraint on the writing modality or content. A set of 12 handwriting indicators, belonging to the temporal, the frequency, the kinematic/dynamic and the non-linear domain, were computed from the raw pen data. The writing signals were acquired from 43 healthy subjects, divided in 3 age groups (young, middle-old, old), while preforming an unconstrained free-writing task.

As first, the reliability of the indicators was checked by calculating the intra-class coefficient (ICC) and the minimum detectable change (MDC) of each indicator in a test-retest trial. The values of ICC, all over 0.75, demonstrated a solid test-retest reliability of the indicators. In the case of changes in the indicators, in longitudinal measurements, the MDC values can help in the discrimination between a measurement error and a variation in the subject's condition. Then, an age-related statistical analysis was performed to study significant differences in the handwriting indicators between differently aged groups of healthy subjects. In this experiment, 8 indicators out of 12 showed significant changes between the 3 groups, with the 95% level of confidence.

Results showed the sensibility of handwriting to variations due to the ageing process. In particular, the older age groups were characterised by an increase of temporal writing measures, a more uniform writing pressure, and more repetitive and predictable tremor oscillation components. Although the actual average value of some indicators could differ from those of previous studies on the same age group, the sensibility of handwriting to age decline was confirmed, even in the semi-uncontrolled setting represented by the free handwriting on paper.

Anomaly Detection in Free Handwriting

The previous investigations revealed that the smart ink pen was a suitable and reliable tool for the ecological collection of quantitative information about the handwriting activity. Although reliable measurements could be obtained using the pen, the subject's handwriting could not be evaluated by simply comparing the indicator with those values which, in previous studies, have been associated to decline. The ecological data acquisition modality increases the noise and the data variability, so measurement may have erratic fluctuations and unforeseen behaviours. In this study, a method based on the anomaly detection was investigated as a solution to identify patterns in the indicators which could suggest an abnormal ageing process or the presence of a pathology. The approach was based on the classification of a subject in a particular group of individuals using his handwriting indicators as input features. By considering groups of differently-aged healthy individuals, an abnormal situation for a subject would be represented by the association of its handwriting features with those of an older aged group. To extend the validity of the study to the more uncontrolled domestic environment, the indicators were calculated from data acquired during writing tasks which simulated the typical handwriting activity in daily life: a short free text and a grocery list, without further constraints.

A total of 80 healthy participants, were equally divided in four groups aged 18-40, 41-60, 61-70 and 70+ (of 20 subjects each). A set of 14 gesture and tremor-related indicators, calculated from the raw data, have been used as features in 5 binary classification tasks between pairs of groups. A baseline classifier (logistic regression) and a more advanced model (Catboost) were used to solve the classification tasks.

Results showed that the Catboost algorithm outperformed the logistic regression in almost all the tasks and the datasets (outcomes ranged 82.5-97.5% for Accuracy, 81.8-100% for Precision, 75-100% for Recall and 92.2-100% for ROC-AUC), in the classifications between groups with close age ranges in particular, where differences in the individuals' handwriting were expected to be minimal. That confirmed the superior sensibility of Catboost to the changes in the handwriting indicators with respect to the baseline estimator.

The objective of these classification tasks was that to test the sensibility of the models to variations in the handwriting due to decline. Especially in the classification between groups of healthy subjects in close age ranges, for which the age-relate differences in handwriting were expected to be limited [16]. For monitoring purpose, Precision was the most significant metrics since it measured how much the classifier was robust in the determination of the true positives. In particular, the classification between the groups aged 61 to 70 and 70+ was the most relevant to study the suitability of the

proposed approach in the early detection of decline scenario.

In a normal ageing process, the signs of decline are expected to be more consistent in the older group [17, 18]. Therefore, whenever an individual in the younger group (aged in the range 61-70) is associated to the older one (70+), it might be interpret as a sign of abnormal decline. In this task, results showed that the trained classifier may be suitable for the decline monitoring application because of its high Precision of 94.4%. As showed in the confusion matrix of the classification between the groups aged 61-70 and 70+, in the upper panel of Fig. 3, only 1 subject to 20 was wrongly classified as older, while the false negative were 3. The ROC curves for this

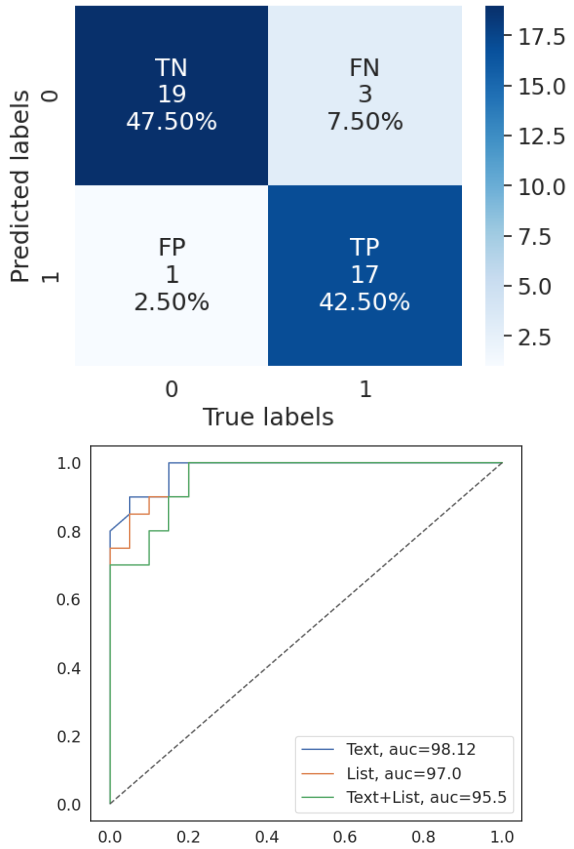

Figure 3: Confusion matrix for the classification between the groups aged 61-70 (labelled as 0) and 70+ (labelled as 1) using the Text data, in the upper panel. The ROC curves for the same classification using the datasets created with the Text, the List and the joined Text and List data are showed in the lower panel.

classification, using the Text and List data, are showed in the lower panel of Fig. 3.

A model explanation technique called SHAP [19] was used to understand the impact that each handwriting indicator had in the determination of the subject's class. Since Catboost was a black box model, the application of SHAP, after the training phase, revealed the importance of each indicator (i.e. how much they weighed in the predictions of the class) and how their high or lower values tended to push the association of a samples towards the younger of the older class. The outcomes showed that the sensitivity of the handwriting features to decline was age-dependent. Moreover, the trend of the indicators in the determination of the subject's age range was analogous with the previous findings in literature: the value of the temporal features suggested that the non-writing moments were longer in the older classes [10], the dynamic features indicated a less fluent handwriting gesture in the seniors and the tremor related features showed a more irregular acceleration pattern in the younger individuals [20].

Whit the same aim of detecting particular handwriting indicators' patterns which could be related to an anomalous ageing process, a population of 20 Parkinson's (PD) patients have been targeted as abnormal group. A similar classification task was then performed between the PD patients (ranging between 55 and 85 years of age) and 20 healthy individuals in the same age range. The Catboost algorithm achieved the best performances in this task: a score of 92% was achieved in Accuracy, 95% in Precision and 90% in Recall. In addition, class wights were used to increase or decrease the importance of a particular class during the training phase. This made the classifier more versatile: for the monitoring purpose, the model was invited to pay more attention to the healthy class, by increasing its weight. It resulted in a maximal Precision (100%), with a Recall of 90%, as showed in the upper panel of Fig 4. On the other side, increasing the weight of the PD class, the Recall reached the maximum score and the Precision resulted equal to 85%, as showed in the lower panel of Fig 4. This made the classifier more suitable for screening applications.

The model explanation SHAP showed that the tremor related indicators had the larger impact in the classification between healthy and PD individuals. Handwriting in PD was characterised by more regular tremor patterns which increased the predictability of the gesture dynamics. The writing force was higher in PDs and the temporal parameters revealed a slower and less fluent writing in patients than controls.

Discussion

The novelty of the proposed solution rely in the autonomous and discreet acquisition modality of the handwriting gesture and tremor data during common paper-and-pen writing activities. The instrument was used to study handwriting in an unprecedented setting where no constraint were imposed

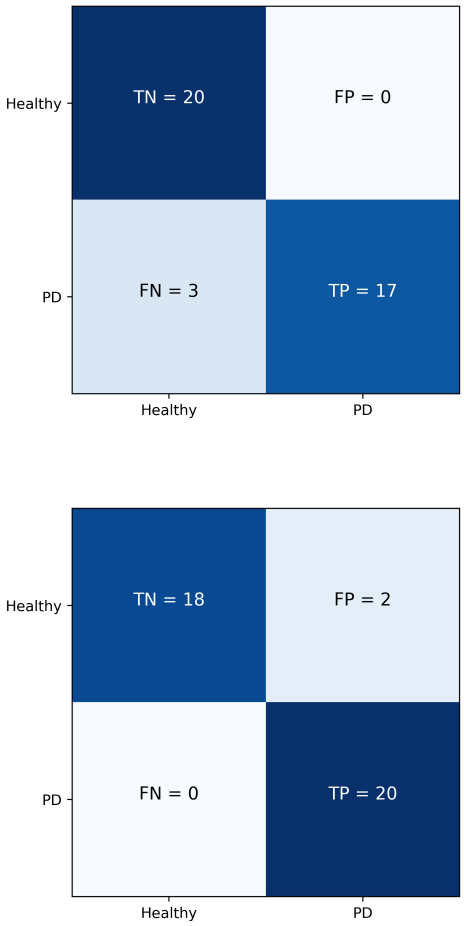

Figure 4: Confusion matrices for the classification between the healthy and the PD groups. In the first case (upper panel) more weight was given to the healthy class. In the second (lower panel) the weight of the PD class was increased.

in the writing exercise. Results showed the reliability of the handwriting indicators in healthy subjects' data and significant differences were found differently aged healthy subjects. The research confirmed previous findings in the literature which, using standard protocols, studied changes in handwriting related to the ageing process. At last, the information retrieved with the handwriting analysis of free text data was used to correctly classify healthy individuals in various age-groups and to discriminate between healthy and pathological individuals. The classification approach was proposed as anomaly detection technique to identify abnormalities in uncon-

trolled handwriting data: the classification of a subject in an older aged group could suggest an abnormal ageing process, while its association with a pathological group might indicate an early stage of the condition. The high values of Precision in the classification tasks showed the low rate of false positive of this supervised and population-specific anomaly detection technique. In addition, an explainable artificial intelligence approach was applied to fully understand the decision of the classification models. The SHAP tool was able to identify the importance of the handwriting indicators and to show how they affected the prediction of the subjects' class.

The proposed solution combined the more traditional signal processing techniques with the novel advancement in artificial intelligence to maximise the information obtainable from noisy and multidimensional data. Raw handwriting data were analysed using conventional techniques to obtain reliable quantities with a precise *physical* meaning. This allowed the association of the handwriting measurements to specific behaviours in the gesture dynamic and tremor domains, and it also include the possibility to clinically interpret the indicators by healthcare professionals. Artificial intelligence was then used to efficiently search in the highly complex space of the handwriting indicators the patterns which may be related to the physical or cognitive decline in the subjects. The effectiveness of this approach is evidenced by the high performances obtained by the classification algorithms, which usually need a large amount of data to match such results. In addition, the use of explainable AI allowed to relate the outcomes of the classification algorithms to the *physically*-explicable domain of the handwriting indicators.

The robustness of the results in this research can be increased by testing the handwriting assessment with a larger amount of data, using the instrumented pen. Moreover, different groups of non-healthy individuals (for example MCI and dementia patients) could be included in the anomaly detection study with the handwriting data, to increase the range of detectable conditions. However, the limitation of this strategy might be related to the use of a population-specific approach. A more sensible anomaly detection method could be studied to monitor the subject-specific changes during time, using longitudinal handwriting data.

A System to Support the Cognitive Assessment

Currently, the detection of the cognitive decline is performed towards the neuropsychological assessment. It consists of a series of various cognitive tests administered by a specialist to evaluate the different cognitive functions. The tests consists of precise clinical protocols which are performed in controlled environments, (i.e. hospitals or clinical facilities) [21]. How-

ever, in the actual setting the diagnosis of cognitive disorders (such as MCI) might happen to be delayed [22]. The main reasons are linked to the facts that i) the examinations are usually prompted by the first manifestation of the symptoms in the individuals and to ii) the long waiting times of outpatient facilities [23]. Moreover, according to a recent investigation [24], the proportion of undetected dementia is above 60%.

Prevention and early detection are envisaged to overcome this issue and lot of research is dedicated to this field. In the eHealth scenario, digital versions of standard cognitive tests have been developed and tested to allow the self-administration in the home environment [25]. The idea is to use these digital tests to send alarm to the senior's family or the general practitioner when the achieved scores are critical.

The Rey-Osterrieth complex figure (ROCF) is a very diffuse test for the diagnosis of MCI and dementia [26]. The *copy* version of this test is performed by asking the subject to copy the ROCF, which is made up by 18 geometrical patters. Then, the clinician evaluates the drawing by singularly scoring each of the patterns. A low total score may suggest some cognitive deficit in the subject.

Its application in a remote monitoring setting could give important information about the cognitive decline in the subject, more efficiently than in the case of an in-person visit. However, the interpretation of the ROCF though the standard procedure may have a poor inter-rater reliability [27]. Although it might not impact the diagnosis in the clinical practice (since a number of various cognitive tests are applied), the low reliability could affect the remote evaluation of the subject and overlook the presence the fleeing signs of a cognitive degeneration [28]. To address this limitation an expert system based on retrospective knowledge have been proposed to support the clinician in the evaluation of the ROCF test.

The objective of this work is to design a decision support system (DSS) with the dual purpose of helping practitioner in the ROCF analysis and allowing the remote evaluation of the self-administered copy-test.

The Decision Support System

The DSS used computer vision and deep learning methods to evaluate the 18 patterns that constitutes the ROCF. The patterns which were composed more by lines or simple shapes were identified in the image with line and shape detection algorithms form CV, then they were evaluated using topological analysis by defining geometrical rules. The more complex-shaped patters were detected and evaluated by estimating a similarity measure with respect to the template image (using a modified version of the ResNet50V2 neural network [29]). The process required an initial action from the special-

ist who was asked to indicate 5 reference points in the image. Then the analysis was completely automated and the algorithms were calibrated/trained using 250 ROCF copies performed from 57 healthy subjects, 131 MCI and 62 dementia patients (from the Instituto Palazzolo, Don Carlo Gnocchi, Milano). The DSS assigned a qualitative label (that could be either *correct*, *misplaced*, *distorted* and *omitted*) to each pattern, according to expert-based rules. Then, it used these labels to predict the most probable diagnosis of the subject between healthy, MCI and dementia. A state of art classification algorithm (Catboost) was used to solve various classification tasks, aimed at investigating the ability of the system to discriminate between healthy and pathological individuals or between a different severity of cognitive decline (MCI and dementia). Binary classification tasks were considered between healthy-MCI, healthy-dementia and MCI-dementia classes, and also a 3-class task including all the groups. Model explanation was also used to understand the impact of the patterns in the classification, as well as their sensitivity to the various classes considered.

Results

The accuracy in the evaluation of the patterns (i.e. in the right assignment of a label among *correct*, *misplaced*, *distorted* and *omitted*) were medium to high: with a minimum value of 62% and a maximum value of 79% of total Accuracy computed as the ration between the number of the correctly and the wrongly labelled pattern.

The classification performance instead were high to excellent: Accuracy scores ranging from 87% (MCI-dementia) to 92% (healthy-dementia); F1 between 79% (MCI-dementia) and 93% (healthy-MCI); Precision between 85% (MCI-dementia) and 100% (healthy-dementia); and Recall between 74% (MCI-dementia) and 90% (healthy-MCI). The lower performance were obtained in the more complex multi-classification tasks: scores from 73% to 76% were obtained for the Accuracy and scores from 73% to 74% for the other metrics. In Fig. 5, the confusion matrices for the classification between the healthy and the MCI groups and between the MCI and the dementia groups are showed. The high performance in these binary classifications showed the good sensibility of the system to different levels of cognitive impairment: when it is in the initial stages with the MCI, and when it become more severe as in dementia.

At last, the model explanation technique SHAP was employed to deeply study the impact of each pattern in the classification tasks. Indeed, the analysis revealed that the sensibility of the patterns changed according to the level of the cognitive function decline and, as expected, the quality of the pattern representation (indicated by the qualitative score) tended to decrease as the severity of the cognitive impairment increased.

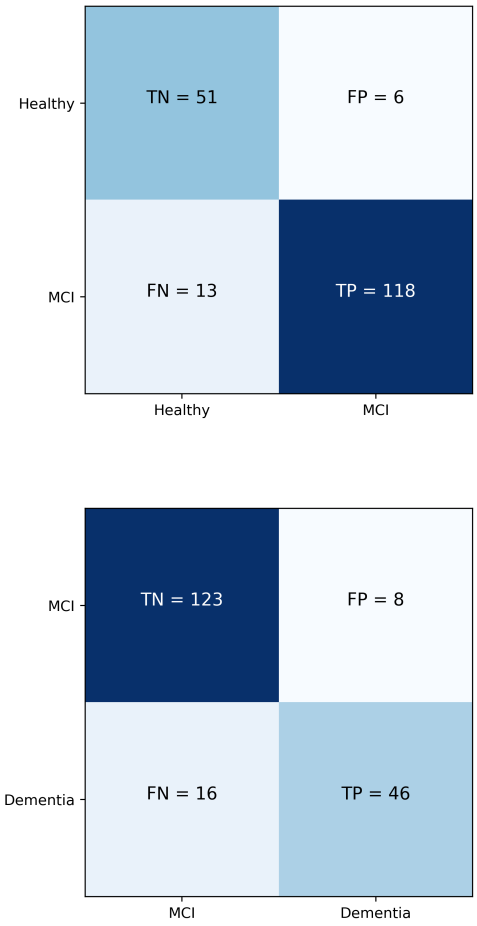

Figure 5: Confusion matrices for the classification between the healthy and the MCI groups, in the upper panel, and between the MCI and the dementia groups, in the lower panel.

Discussion

The presented DSS used computer vision and deep learning algorithms to detect and evaluate the 18 patterns in the ROCF. Then, the scores assigned to each pattern were used to classify the sample with the most probable diagnosis with high levels of performance. The results demonstrated the feasibility of an expert system which performs a semi automated analysis of the ROCF. Domain experts could use the system to have suggestions about the evaluation of the single patterns of the ROCF and the most probable diagnosis for the examined individual. The application of explainable AI can

also supply further information on how each pattern of the ROCF copy have effected the final decision. Furthermore, the DSS allows the remote (and posterior) analysis of the ROCF copy test. In remote monitoring scenario, the expert system could be used by the practitioners to evaluate the ROCF drawings done by subjects during tele-consultations or in self-administered tests. Poor test results may suggest a possible degradation of cognitive function.

The DSS outcomes are based on retrospective knowledge, therefore it will improve in accuracy and robustness as the number of data concerning past examples increases. The limited number of samples in this study might have affected the accuracy in the evaluation of some patterns, yet the performance in the classification of the diagnosis were surprisingly high. Especially considering that the ROCF copy test is just a part of the cognitive assessment procedure.

A very large collection of ROCF copies could improve the performance of the DSS in the pattern evaluation and increase the robustness in the classification of the pathology. Higher amount of data could also encourage the use or more unsupervised approaches in the analysis of the figure, which it may exclude any subjective component in the evaluation of the ROCF.

Conclusions

This PhD research proposed novel eHealth methodologies to support the early identification of age-related decline. Remote monitoring of community dwelling elders and the remote assessment of the physical or cognitive functionality are possible strategies to improve the probability to early diagnose a chronic disease. Therefore, this work proposed a solution for the remote monitoring of the older adults physical and cognitive decline through the ecological assessment of handwriting and an expert system to help the specialists in the remote examination of the Rey-Osterrieth complex figure (ROCF) test, which is an important tool for the diagnosis of MCI and dementia.

The first objective consisted in a transnational study in which the application of the handwriting analysis, as biomarker for the age-related decline, was investigated and adapted in the unprecedented context of remote monitoring. Being a common exercise and having been observed to vary with ageing and pathology, handwriting appeared as an optimal candidate for the instrumental daily-life activity monitoring purpose. In literature, the handwriting assessment consisted in the analysis of certain indicators, computed from the signal acquired during the handwriting activity. These quantities, related to the temporal, dynamic and tremor characteristics of

the writing gesture, were studied to undergo variations with the presence of age or pathological decline in the individuals. However, some limitations were found for the application of the handwriting assessment in the remote monitoring in the home environment. A first barrier was represented by the data acquisition instruments (tablets or digitizers), which were not easy to be independently operated by older users. Then, the use of standard protocols for the data acquisition in literature did not allow to extend the results of the previous studies in the uncontrolled domestic environment.

To overcome these limitations, a sensorized IoT ink pen was developed to allow the ecological handwriting gesture data acquisition in paper-and-pen tasks. The device was successfully tested, validated and then used to collect data from healthy subjects. Then, data were used to study the handwriting assessment in a conditions which was similar to the uncontrolled remote setting. As first, the reliability of the indicators was confirmed in test-retest writing tasks. Then, their sensibility to age-related variation was studied in 3 differently aged groups of healthy subjects. In a total of 12 indicators, 8 of them showed significant changes with age. This result confirmed the possibility to use the handwriting assessment as an instrument to detect variation related to the ageing process in uncontrolled environments. Although the variation trend of each indicator with age was analogous with those reported in literature, their mean values not always corresponded with the previous studies in controlled settings. As a consequence, a remote monitoring system based on the handwriting assessment should not rely on standard reference values for the indicators, but it should consider a multivariate approach in which various different patterns of indicators may be associated to age or pathological variations in the subject.

A supervised and population-specific anomaly detection approach was presented as a method to exploit the handwriting assessment in the remote monitoring of the subject's age or pathological-related decline. The aim was that to use the handwriting indicators to classify subjects in the correct age-class or pathological (PD) group they belonged. In the remote monitoring context, the unexpected classification of an individual (for example a subject aged between 60-70 classified as an over 70 years of age, or an healthy subjects classified as a PD patient) may indicate a sign of an abnormal ageing process. Therefore, such event should trigger a deeper investigation on his health status. This strategy was tested by solving various classification tasks between group of differently aged healthy individuals and a group of PD patients. A state-of-art classification algorithm (Catboost) was used and the achieved performance were high in terms of Precision and Recall in all the tasks. Since the classification algorithm was a black box model, a recent explanation technique (SHAP) was applied to understand each model decision. This additional method increased the interpretation of the anomaly detection strategy, as it revealed the impact and the behaviour of the handwriting indicators (known quantities) in the identification of the subject's

group.

A critical point of this proposed solution can be found in the limited amount of data that could be collected (80 healthy subjects and 20 PD patients). Although the modest number of samples, very good outcomes in terms of classification performances were obtained. Yet, a more robust performance estimation would be envisaged using a larger number of individuals. Furthermore, the addition of more pathological categories, (such as MCI and dementia), would enlarge the spectrum of the abnormal detectable conditions. Another limitation can be addressed to the population-specific nature of the anomaly detection methods which was tested in this research. A more subject-centred approach would increase the sensibility in the identification of the abnormalities in the handwriting data, overcoming the problems related to the inter-subjects variability. However, longitudinal studies are necessary to acquire subject-specific data which may supply information about the individual's decline trajectory.

The second objective of this PhD work consisted in the design and development of a decision support system (DSS) for the analysis of the Rey-Osterrieth complex figure (ROCF) copy test. To overcome the limitation in the efficacy of the spot visits in the early diagnosis of cognitive decline, various eHealth strategies have been investigated in literature to allow the remote administration or evaluation of some cognitive tests: they can be remotely supervised during tele-consultations or self-administered by the subject and forwarded to the specialist.

The application of the ROCF test in a remote monitoring setting could give important information about the cognitive decline in the subject, because of its clinical relevance. However, some studies pointed out that the interpretation of the test though the standard procedure may have a poor inter-rater reliability. Although it might not impact the diagnosis in the clinical practice (since a number of various cognitive tests are applied), this issue could affect the remote evaluation of a self-administered test by the subject and overlook the presence of fleeing signs related to the cognitive degeneration during time. To address this limitation an expert system have been proposed to support the clinicians in the evaluation of the ROCF copy test.

The DSS was based on retrospective knowledge collected from past ROCF copies examples from normal, MCI and dementia individuals. It used computer vision and deep learning algorithms to detect and evaluate the 18 patterns in the ROCF, by assigning a qualitative score each. Then, it used the scores to classify the image with the most probable diagnosis. The DSS was able to correctly discriminate the ROCF copy tests between healthy and MCI individuals, and between MCI and dementia patients, with good levels of Precision. Domain specialist could use the system to have suggestions

about the evaluation of the single patterns of the ROCF and about the most probable diagnosis for the examined individual, especially in a remote assessment framework. In addition, the information the specialists would retrieve by using the DSS was enriched with the application of the explainable artificial itelligence (AI) model SHAP. This tool returned the impact of each of the 18 ROCF's patterns in the classification outcomes, so it revealed how the sensibility of the patterns changed with the progression of the cognitive decline in the individuals. The system could be also used in research, as it might highlight correlations of different mental disorders or impaired functionalities with particular elements in the ROCF copy test.

A grater number of samples could improve the accuracy of the DSS in the pattern evaluation and enhance the robustness in the diagnosis formulation. A larger dataset could also open the way to the exploitation of unsupervised deep learning approaches which may exclude any subjective component in the evaluation of patterns. Therefore, more reliable suggestions would be supplied to the specialists using the DSS.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Scientific Context

Demographic ageing is a global trend. In the European Union (EU), the number of people aged 65+ will almost double over the next 50 years. While increased longevity is a great achievement, it is also a formidable challenge for health and social care sustainability. The main reason is that ageing is generally accompanied by a progressive decline in physical and cognitive function that in some cases may develop into, or lay the foundation for, one or more chronic conditions. Neurodegenerative disorders such as Alzheimer's and Parkinson's diseases have a high prevalence in individuals aged 65+ (7% and 1% respectively) and it increases with age [30, 31]. Moreover, other conditions are widespread in the elderly population. Among them, frailty have the largest incidence in older adults and it leads to a spiralling decline in various functional domains that aggravates the risk of geriatric syndromes [4]. The management of chronic diseases demands a huge use of resources by both patient's families and health systems, as well as leading to a poor quality of life for the individual itself. The degeneration process in the vast majority of age-related syndromes is irreversible. However, the decline can be slowed and the negative effects can be mitigated if interventions are timely provided. In the early stages, the conditions assume a moderate form with limited effects on the individual's physical and mental functionality. For example, mental decline starts with a mild cognitive impairment and frailty is anticipated by the pre-frail condition. In these precursory phases, symptoms have a minor impact in the functional ability of the patient. Yet their identification is crucial for the well-timed detection of the condition.

Early detection of decline is decisive to prompt the appropriate intervention. However ageing is a complex and multi-factorial process and the course of decline is very subject-specific. When in their early stages, the signs of some age-associated pathological conditions may be hidden by the

typical manifestations of normal ageing. In addition the current healthcare model, which almost rely on spot-visits, and the lack of uniformly accepted standard tools for the effective early diagnosis make the timely detection of decline even more arduous.

In this scenario, the new requirements for the healthcare system envisage a shift towards the prevention and the early diagnosis of chronic disease in older adults. Technologies such as eHealth and internet-of-things (IoT) can help this transformation by enabling remote home monitoring solutions. The possibility to remotely track the health status in their home has various benefits. The continuous monitoring can allow the timely detection of significant variations due to decline and it may optimise the access to the hospital resources by avoiding unnecessary visits and hospitalisations. Remote monitoring also supports ageing in place, i.e. it allows seniors to safely spend their late years within their habitations and communities.

Home monitoring research is still underway. The major issues are related to the uncontrolled setting in which the measurements are performed. Therefore, standard metrics are hardly identifiable. Then, with older adults, the complications increases because the adoption and the acceptance of the monitoring devices is prevented by the widely diffuse senior's technological illiteracy. For the remote tracking of the physical functionality, instrumental daily-life activity monitoring is a solution which may results non-intrusive (or ecological) for the user, since it basically collect data by carrying out daily activities. They usually involve the use of a sensorized common object which does not request the users to adopt new behaviours or to perform tasks that may results uncomfortable. The home-based cognitive monitoring instead is more related to the use of standard protocols, which may be presented in the form of smartphone and tablet applications or serious games. The validity of some self-administered digital cognitive tests for the older adults have been investigated in literature [25]. The score the older adults receive in those tests supply useful information about their mental decline and it can be timely forwarded to the clinician. The major issues in this field concern the implementation of automatic method for the test evaluation and its association with cognitive decline.

The general aim of this work is expressed as follows: *this research has the objective of designing and developing leading edge eHealth technologies and methods to support the detection of physiological changes in healthy ageing and to detect signs of pathological decline.*

This study is mainly focused on the exploitation of the ecological assessment of handwriting as daily life activity monitoring. Handwriting has been largely investigated in literature to be a potential biomarker for both physical and cognitive decline, since it is an high level skill which involves several

cognitive and motor systems in the individual. Pathological decline, due for example to a neurodegenerative process, affect some temporal and frequency characteristics of the writing gesture and variations in quantitative indicators of handwriting are observable. Moreover, the measurable features are also altered by the physiological ageing course. That means that abnormal variation in an older individual may suggest the onset of some non-healthy decline. However, there are limitations to the application of the handwriting assessment in domestic environments. As first, almost all the scientific findings in literature were achieved in controlled settings, where standard experimental protocols were adopted. This made the translation of the previous results in the new uncontrolled context of the home monitoring not straightforward. Then, the available instruments for the handwriting data acquisition (tablets and digitizing surfaces) may not be easy to be independently used by seniors. Therefore, a handwriting monitoring system based of such technologies can result too intrusive to be adopted in the daily-life.

As second task, this work deals with the problem of the early detection of mental decline through the use of digital cognitive tests. In the clinical practice, the cognitive assessment rely of the administration of several tests and the outcome is determined by the clinician according to the score the patient achieved in each of them. Yet the self-administration of some tests by the subject does not replace the comprehensive clinical cognitive assessment, but it might serve as sign of early decline in the case of a scarce outcome. The Rey-Osterrieth complex figure is one of the most informative cognitive tests, since it evaluates several mental functionalities. However, its interpretation is a complex process which can not be easily automated by a computer software. Furthermore, its evaluation does not achieve a high level of inter-rater reliability [27]. Part of the information indeed rely on the subjective experience in evaluating also the patient's execution strategy of the test. For this reason, the remote posterior analysis of a self-administered Rey-Osterrieth test is almost impractical. This study proposes a decision support system for the interpretation and the evaluation of the Rey-Osterrieth complex figure. It is meant to help clinician in the evaluation of the test and to also to allow the remote analysis of a self-administered test, performed in the home environment.

1.2 Aim

This work aims at designing and developing novel technologies and methods to support the detection of age-related decline. Two approaches are here proposed: *the ecological assessment of handwriting as instrumental daily life activity monitoring* and a *decision support system to assists the clinicians in the analysis of the Rey-Osterrieth complex figure copy test and also allow its remote evaluation.* The objectives of this research consist in solving the

scientific and technological problems for their realisation.

In particular, for the ecological monitoring of handwriting they are:

- 1a) the lack of an instrument which allow the non-intrusive collection of the handwriting data;
- 1b) the lack of scientific evidences which demonstrate the validity of handwriting as biomarker for age and pathological-related variations in uncontrolled settings;
- 1c) the lack of verified methodologies to identify these variations in uncontrolled settings.

While, for the implementation of a decision support system for the analysis of the Rey-Osterrieth complex figure, the issues are related to:

- 2a) the complexity of the figure to be analysed;
- 2b) the difficulties in detecting and quantify the aspects not included in the standard evaluation protocol, but still present in the analysis of the test.

Consequently, the following solutions have been addressed to resolve the respective problems. For the first objective they are:

- 1a) design and validation of an instrumented ink pen, to allow handwriting data acquisition during common paper-and-pen tasks;
- 1b) study the validity and the reliability of the main handwriting indicators as biomarker of decline in semi-uncontrolled settings;
- 1c) the investigation of anomaly detection techniques based on the discrimination between age-groups of healthy subjects and between age-matched groups of healthy and pathological individuals.

Then, for the second objective they are:

- 2a) the application of cutting edge computer vision and deep learning algorithm for image elaboration;
- 2b) the use of retrospective knowledge on past Rey-Osterrieth tests to determine the most probable outcome.

1.3 Outline

The rest of this thesis in organised in 6 chapters.

- Chapter 2 introduces the context and the state of the art related to this work. It regards various subjects as the phenomenon of the demographic change towards the older age, the physiological and economical issues related to ageing, the new requirements for the healthcare systems and the home monitoring framework are treated.
- Chapter 3 presents the scientific literature about the quantitative analysis of handwriting and its limitations for its ecological application.
- Chapter 4 describes the design and the development of the smart ink pen. The calculation and the validation of the handwriting indicators and the study of their differences between age-groups of healthy individuals. The last section of the chapter reports the results of a ecological momentary assessment study about the usability of the pen among seniors.
- Chapter 5 is dedicated to the topic of anomaly detection of decline through the group-classification based on the handwriting indicators.
- Chapter 6 explains the implementation of the decision support system for the evaluation of the Rey-Osterrieth complex figure test.
- Chapter 7 at last draws the conclusions of the work.

Chapter 2

Background Information

This chapter will treat the general aspect about ageing. Sec. 2.1 introduces the topic of ageing from a demographic and physiological point of view. Particular attention is given to the typical trajectories of decline in elders and the importance of early detection of chronic diseases. Sec. 2.2 explains the impact of ageing to the healthcare systems, highlighting the actual limitation and discussing how the healthcare model should be changed to efficiently face the phenomenon of the ageing population. A brief section is also dedicated to the issues of the healthcare systems revealed by the current COVID-19 pandemic. Sec. 2.3 shows home monitoring as a solution to improve the elderly care efficacy and to offer the possibility to early detect sign of decline in community dwelling elders, with a section dedicated to its design specification guidelines and requirements.

2.1 Ageing

Nowadays, the increase of the life expectancy in high-income countries is allowing life beyond 60s for most of the people [32]. The combined effect of a reduced mortality among older adults and a significant decline in fertility is determining an unprecedented modification on the composition of the population that will drive a demographic shift towards the older age [33]. These demographic projections are widely predictable, so institutions can plan decisive actions to face the deep impact of the next transformation of our society [34]. *From the family level to the local community, older people can offer a valuable contribution to the society, as long as they live in good health. However, if the extra years are lead by physical and mental decline, the effects may be counteractive for the society and for the elder individuals themselves.*

As evidences suggest, increased longevity does not always mean good health. Longitudinal studies in wealthy countries observed that the prevalence of more severe disabilities which appeared decreased, while the preva-

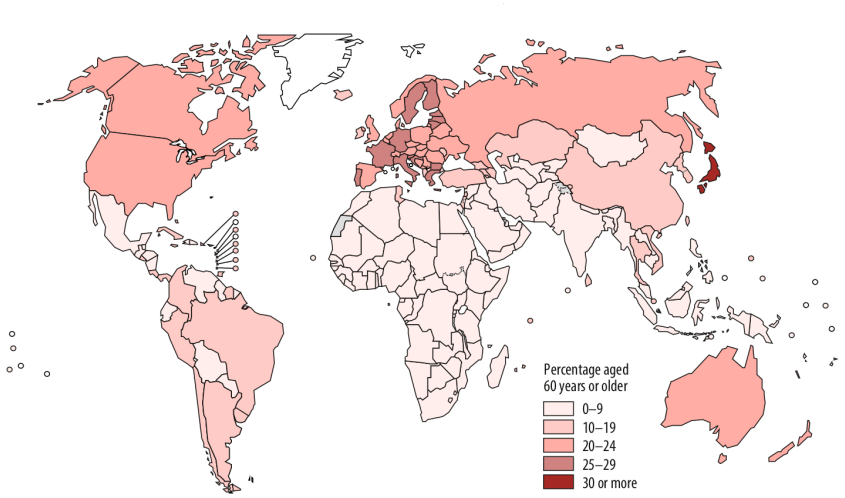

Figure 2.1: Proportion of population aged 60 years or older, by country, 2015, from https://www.who.int/ageing/publications/world-report-2015/en/.

lence of low severe disability remained almost unchanged [35, 36, 37]. The health problems of older people are mostly determined by chronic disease. Therefore, many of these can be prevented or slowed down with the assumption of healthy behaviours and early diagnosis. Physical activity, good nutrition and the avoidance of harmful behaviours can promote physical and mental well being and the early detection of health problems can allow an efficient management of health. Elder people can still conduct a dignified life and they can contribute positively to society, if the long-term care are able to provide the right support [1]. Hence, The phenomenon of the ageing population needs an appropriate response by public health systems.

2.1.1 Demography of Ageing

At present, the ageing population is a worldwide fact and a key policy issue for the governments of all countries. The portion of aged people is drastically increasing all over the world. In 2015, the proportion of population aged 60 years or older, by country, is reported in Fig. 2.1. The map shows that in 2015 only Japan exceeded a portion of older population of 30% [1]. Projections, nevertheless, show that similar demographic characteristics will be shared among many other countries. Figure 2.2 shows the projections of the proportion of people aged 60 years or older by country for 2050 [1]. Countries in Europe and North America are among them, but also countries such as China, Iran, the Republic of Korea, the Russian Federation, Thailand and Vietnam. Beside the global increasing life expectancy, the ageing population in some countries is also due to the improve survival at younger ages that took place within the socioeconomic development during the past

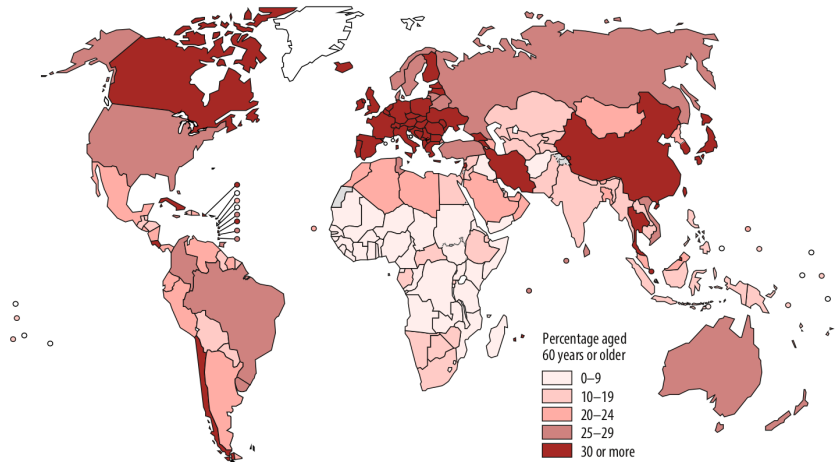

Figure 2.2: Proportion of population aged 60 years or older, by country, 2050 projections, from https://www.who.int/ageing/publications/world-report-2015/en/.

50 years.

However, the increase of life expectancy and the new social transformations may result in an incremented gap between different co-living generations. Today it is not so rare that families include members belonging to different generations, but they live separately with more probability than they did in the past. As an example in the European Union, the portion of 65 and older aged women living alone is more than 40% [38]. For older people living alone, isolation can be an issue as it increase the probability of undesired outcomes [39]. In these cases, remote monitoring devices and assistive technologies may help ensure a safer lifestyle for community dwelling elders. Together with social inclusion policies and activities [1].

2.1.2 The Process of Ageing

Biologically, the process of ageing consists in a lifelong accumulation of molecular and cellular damage. Over the years, these phenomena lead to a progressive and generalised decay of physiological functions, extreme sensitivity to environmental factors and a growing risk of developing a pathological condition and death [40]. The phenotype of the decline course is various and it usually involves several body functions.

Motor functions. The neuro-motor system undergoes major changes during the decline process. First, the muscle mass tends to progressively decrease causing a reduction in strength and musculoskeletal functions [41]. Grip strength, for example, is one of the indicators used to monitor the motor-system ability to generate force, regardless the effect of any possibly present

disease. Furthermore, strength decline has also been studied to be a strong predictor of mortality [42]. Mobility is also impaired by important changes in bones and articulations. The bone density tends to decrease with age up to causing the condition of osteoporosis, in which the possibility of suffering a fracture is greatly increased [43]. Articulations become more rigid and fragile with age, since the tissues as joint cartilage undergo cellular and molecular deterioration [44]. From a macroscopic level, all those changes and degradation in other factors, such as coordination and proprioception, impact higher musculoskeletal functions. The reflections can be seen in a reduced gait velocity, a reduced speed and accuracy in performing daily-life motor tasks and an overall reduced mobility [45]. These physiological modification generate different mechanism which can favourite the onset of neuromotor disorders. The potential contributions to age-related neuromotor disorders, such as Parkinson's disease (PD), essential tremor (ES), multiple sclerosis (MS) and stroke, are multi-factorial and they can include a large number of conditions, drugs and environmental hazards, very often in combination [46].

Sensory functions. Ageing is generally accompanied in a degradation of the sensory functions. With a large diversity at the individual level, impaired vision and bilateral hearing is experienced in older adults. Presbyopia is the most common condition associated to complex functional changes in the eye which lead to a decrease of focusing ability [47]. If not treated, these physiological modifications can affect fundamental aspects of daily life of elders by compromising the normal performance of daily activities. Hearing loss can also place barriers to communication, favouring conditions such as social isolation, loss of autonomy, depression and cognitive decline [48]. A reduced vision also contribute to isolation by making interpersonal interactions and the access to information more complicated. The generalised loss of sensory function eventually affects mobility and the execution of habitual activities, increasing the risk of falls and accidents.

Cognitive functions. The decline of cognitive functions is another important aspect of ageing, which may occur in relatively young age. The cognitive decline is a complex and dynamical process that evolves in different conditions at different rates. With the increasing of individual's age, the range of possible impairments of cognitive functions becomes more and more heterogeneous [49]. Cognitive decline is commonly associated to a reduced ability to perform complex tasks which demand divided or selective attention [50]. Learning and complex memory tasks which require reorganisation and integration of several items indeed become limited with ageing. However, memories related to personal experiences and procedural memory remain usually intact. This means that, from a cognitive side, high level learned skills such as riding a bike remain almost stable in physiological ageing [51]. Age-related cognitive decline can be slowed and compensated, to some extent, with mental training and physical activity [52]. Pathological

cognitive decline instead might be harder to reduce. Early detection and intervention are key to slow the decline process which usually starts with moderate forms, such as in the mild cognitive impairment (MCI) condition [53], and progresses to more severe forms of dementia, as the Alzheimer's disease (AD). Cognitive impairments have heterogeneous manifestations in older adults, but normal ageing decline is still distinguishable from changes due to pathological conditions [16].

Dementia is not an inevitable part of ageing [54], however it does have a very large incidence. The people affected by dementia in 2015 were more than 47 million worldwide and projection are less comfortable: more than 75 million people are estimated to have dementia by 2030 [1]. Dementia is definitely one of the today's and next future challenges because of its huge personal, social and also economic consequences.

2.1.3 Co-morbidity and Frailty

The process of ageing is multidimensional and the typical effect is a gradual drop in the capacity of the individual [40]. The course of this decline is non-linear and not consistent among different subjects [55]. Some older individuals can keep good physical and cognitive functionalities, while other can live in a condition of non self-sufficiency. At the basis of this phenomenon there is a component of variability given by the complexity of the process, but a strong influence is also given by environmental and behavioural factors.

At any rate, with ageing the risk of developing chronic diseases increases and often they appear in the form of noncommunicable disease. By the age of 60, the risk of multi-morbidity is augmented, that is the presence of more than one chronic or acute condition at the same time [56]. When an elderly individual experiences the coexistence of multiple diseases, these can determine health conditions that might escape the standard methods of detection. Moreover, they should not be treated separately. A comprehensive health assessment of an older individual should consider the mutual interaction that the various morbidity can have and how they influence the possible trajectories of physical and mental functioning [57].

Sates of comorbidity can emerge in the chronic condition of *frailty*. A standard definition of frailty is not yet uniformly recognised, however it can be described as a progressive decline in physical and/or mental abilities which leads to an accentuated vulnerability to stressors and tendency to more adverse health outcomes [58].

The prevalence of frailty in Europe, in people aged between 50 and 64, is massive. It is estimated to be 4.1% and it increases up to 17% in individuals aged 65 and over [59]. Furthermore, the prevalence values of the precocious state of frailty, the pre-frailty, are significantly higher: 37.4% for people aged

within the range 50 - 64 and 42.3% for those aged 65 and over. Values also resulted notably higher in southern Europe.

Frailty progresses as a complex and dynamical process in the individual [60]. The course can vary greatly from person to person. In some cases it appears reversible, but only in rare cases the process regresses spontaneously [61]. In most cases only a well-timed and comprehensive geriatric assessment can provide indications for administering effective interventions aimed at preventing major negative outcomes [62]. However, the identification of frailty at its early stages might be complicated as symptoms can be analogous to those typical of normal ageing [3]. Fried an colleagues described the generic phenotype of frailty [4]. They identified five typical signs related to a non-physiological ageing process:

1. *Shrinking*, detectable by an unintentional weight loss equal or higher 10 pounds in prior year.
2. *Weakness*, when the measured baseline of grip strength is lower than 20% of the standard value, adjusted for gender and body mass.
3. *Poor endurance*, as self-reported exhaustion associated with stage of exercise reached in graded exercise testing.
4. *Slowness*, based of standard time to walk a distance of 15 feet, adjusted for gender and height.
5. *Low physical activity*, identified by the lowest quintile of physical activity in terms of a weighted score of kilocalories expended per week.

According to Fried, the presence of just three of those signs in an older individual indicates frailty. The manifestation of two symptoms only instead, may imply the pre-frailty state in the individual.

Another instrument which has been proposed to detect frailty is the frailty index [63, 64]. It is a score, continuous between 0 and 1, which counts the health deficits of an individual and divides that amount by the total number, at list 30, of deficit measured. The health deficits can be either signs, symptoms, diseases, disabilities and abnormal diagnostic test results. Moreover, to be valid , the health deficits should involve multiple domains of functioning or multiple organ systems, the prevalence should rise with age, they should be not so frequent before the age 65 and the prevalence should be over the 1%. Higher values of the frailty index indicates higher severity of frailty, but the cut-off is questionable. In most cases a value greater than 25 is considered as threshold for frailty.

This lack of a standard tool to identify frailty makes sporadic visits and assessments poorly efficient in the detection of the syndrome. Longitudinal research on trends and trajectories of frailty are therefore of high priority in research, as well as randomised controlled trials pointed on prevention or treatment of frailty [65].

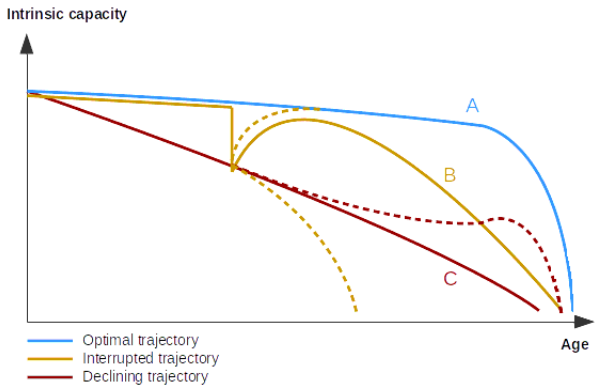

Figure 2.3: Three hypothetical trajectories of the intrinsic capacity, adapted from https://www.who.int/ageing/publications/world-report-2015/en/.

2.1.4 Trajectories of Decline in the Elderly

In the sociopolitical and in the research environments, *healthy ageing* is a rather widespread term. It indicates the good health status of an individual in older age as opposed to the presence of single or multiple diseases. However its exact definition as well as an adequate instrument to measure it is still lacking [66]. According to the World Health Organisation's *World Report on Ageing and Health*, of 2015 [1], healthy ageing can be defined as the ability of elders to develop and keep the functional ability that enables well-being. Functional ability is a name used to express health characteristics that allow individuals to perform normally expected activities and tasks. It is composed by the *intrinsic capacity*, which includes all the physical and cognitive capacities of a person, environmental factors and the interaction of the individual with them. Well-being is often assumed in a broad sense, comprising abstract characteristics such as happiness, satisfaction and fulfilment.

During the ageing process, the way individuals interact with the environment start to change from a certain point in midlife. The way these interactions change can draw different trajectories in the intrinsic capacity and in the functional ability, according to the various health conditions people can fall in. Figure 2.3 shows three hypothetical ways the trajectory of the physical capacity can evolve. The trajectories start from the same point on the time and the event of death is assumed to happen nearly at the same age. What is shown to change is the amount of life lived with an high level of intrinsic capacity, which positively contributes to the quality of life. The blue line, case (A), describes the trajectory of an optimal ageing course in which the intrinsic capacity rapidly falls starting from a point close to the end of life. The yellow line, case (B), represents the trajectory of an individual which had a rapid decrease in the intrinsic capacities caused by an event

such as an acute condition. The rapid decrease is followed by a recovery tract, which then turns into a progressive decline. The red line, case (C), shows the trajectory of a hypothetical subject with a chronic condition, in which decline constantly progresses until death. The dashed lines describes alternative evolution for the trajectories in the cases B and C. Before decline has deflected the trend of the intrinsic capacity of an individual, the objective for an healthy ageing is that to favourite an optimal trajectory, as in the case A.

Evidences show trajectories of intrinsic capacity can be monitored and used to predict the future progressions [67]. Besides health characteristics of the subject, other information are needed to improve the predictions, such as personal habits, measures about daily-life activities and genes factors. In the case where subjects are living with chronic diseases, intrinsic capacity can be improved if disease is promptly detected. For the Alzheimer's disease (AD) for example, as reported by Rasmussen and Langerman in 2019 [2], early detection positively impact the individuals in terms of quality of life, as they are more likely to control their conditions and live independently for longer. A better quality of life can be maintained for several years. Timely treated AD patients showed were associated to a reduced tendency of death and a lower rate of hospitalisation, respectively the 31% and the 20% less than non-treated patients [68]. Early diagnosis of cognitive decline also have positive impact on the economical burden for long-term care systems. The effect of medical treatments at early onset of a chronic disease have the effect to slow its progression and stop, in some cases, the development of more severe conditions.

Early intervention also results in a reduced expense or long-term care systems. The graph in Figure 2.4 shows the projections of total medical and long-term care costs in case of the current diagnosis status (in violet) and in case of early diagnosis (in green), up to 2050. The projections were made by the Alzheimer's Association using the *The Health Economic Medical Innovation Simulation* model [69]. According to the model, partial early diagnosis refers to the situation in which people are more likely to receive an AD diagnosis during the MCI stage rather than the dementia stage. The prediction showed that, in the case of early diagnosed disease, savings will be retrieved and estimated to increase about the the 13% in 25 years.

2.2 The Healthcare Model

As outlined by the World Report on Ageing and Health [1], today healthcare systems are oriented at the diagnosis and treatment of health issues limited in time. The common approach consists in searching the problem and act the clinical procedures to cure it. The result is a healthcare model which is

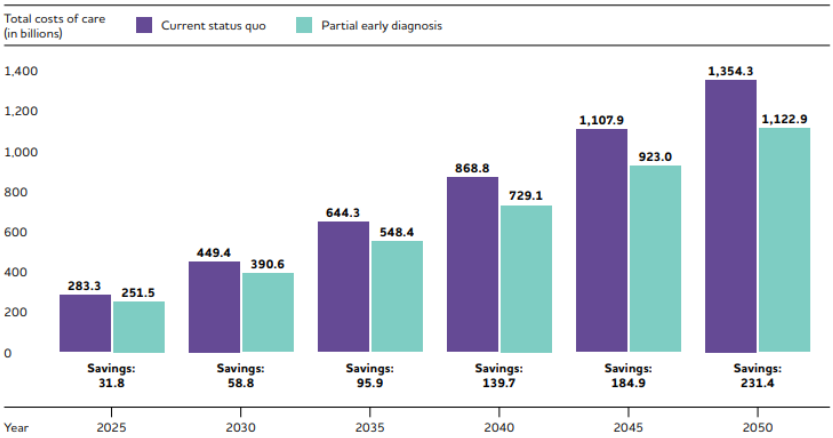

Figure 2.4: Projected total medical and long-term care costs and cost savings by diagnosis scenario, from https://www.who.int/ageing/publications/world-report-2015/en/.

optimised to tackle acute conditions and communicable diseases. Long-term health has always been addressed with lower priority by healthcare policies, at the expense of the older individuals' needs.

Factors that require long-term care increase with age, as noncommunicable diseases become more frequent ad comorbidity is more widespread. In addition, to better address the needs of elderly care, a high degree of coordination and cooperation between several health professionals would be requested. Currently, while the demand of healthcare for the elderly population is increasing, healthcare systems are not yet sufficiently suited to facilitate such coordination, therefore, the delicate task of communicating relevant to their health situations is left to the same individuals and their families.

The formation of health practitioner should be also revisited to efficiently include the long-term care in the healthcare model. Developed to deal with the wide spread acute infectious diseases of the 20th century, the training of practitioners was mainly focused on identifying and treating the symptoms with an episodic approach [70]. Elderly care, on the other hand, would benefit from the tendency of practitioners to anticipate and counteract the deterioration of the patient's functional status [71]. Such attitude would better fit the comprehensive characteristic of a more older-suited care [72, 73].

2.2.1 Economic Impact

Today, healthcare for the elderly already has a considerable impact the public healthcare spending. The Ageing Working Group of the Economic Policy Committee (AWG) constantly monitors the public expenditure on both

health and long-term care (LTC) by making predictions using the European Commission services' models (EC and EPC, 2017).

Figure 2.5 shows the LTC spending as a percentage of GDP for the European countries reported in the OECD/EU document *Health at a Glance: Europe* 2018 [74]. Long term care expenditure emerges as a growing part of GDP in many EU countries and, therefore, it had a large impact in the long-term sustainability of public finances. Projections under the AWG setting estimate a baseline scenario of The LTC spending in European countries raised of more than one percentage point, from 1.6% of GDP in 2016 to 2.7% of GDP in 2070. Such scenario can widely vary across the 28 EU countries, from only 0.1 percentage point of GDP in Greece and Bulgaria up to more than 2 percentage points of GDP in Luxembourg, Netherlands and Denmark.

As showed in Section 2.1.4, the health status of an ageing individual can follow different trajectories. Therefore, it turns out that time to death is a more reasonable predictor of the healthcare costs than age [75]. This is partly due to the complexity of the care contexts that the elderly need, but also to the structural and cultural attributes of health systems and policies. *For instance, low effort is dedicated to prevent hospitalisation, which is very expensive and necessary when chronic diseases progress to the more severe states* [76].

Today, the worldwide rapid increase of the elderly population places at the center of the debate an appropriate revision of the healthcare model. A different and elderly-centred approach is envisaged to foster the so called *healthy ageing*. The goal should be to maintain, as much as possible, a high level of intrinsic capacity in older adults who experience healthy conditions and to prevent disease and the risk of negative outcomes. *Special emphasis should be given in detecting and managing noncommunicable conditions at early stage*, without disfavouring adequate interventions in the case of acute

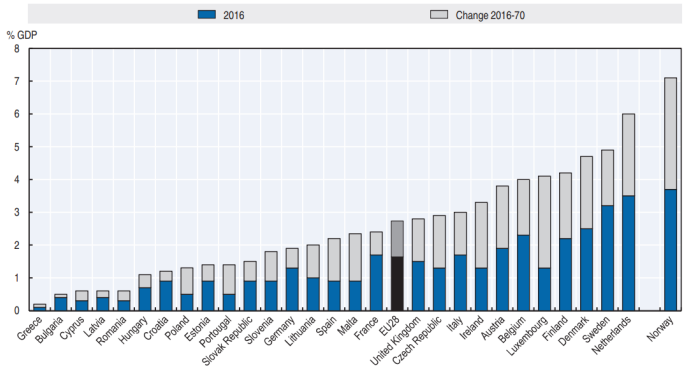

Figure 2.5: Public spending on long-term care as a percentage of GDP, 2016 to 2070, from http://dx.doi.org/10.1787/888933837036

problems [77, 78].

2.2.2 The COVID-19 Impact

The ongoing COVID-19 pandemic has revealed and emphasised the vulnerabilities of health systems worldwide with brutal strength. Since the beginning of the large diffusion of the SARS-CoV-2 virus in early 2020, frequent hospital visits are being discouraged. The massive use of hospital resources and structures for the intensive care and the need to limit the risk of contagion resulted in a significant decline in common healthcare services. A recent review of the data relating to 20 countries, including Italy, highlighted an overall reduction of 37% in healthcare services, higher for outpatient visits (42%) and lower for hospitalisations (28%) , diagnostics (31%) and therapeutic treatments (30%) [79]. Consequences on people's health are critical. There are increasingly multiple accounts of inconvenience, distress and adverse events attributable to delays, barriers and deferrals of essential services that affect different health sectors.

In order to limit health worker exposure to ill patients, conserve personal protective equipment and decrease the effect of patient surges on hospitals, improvements in the way health care is provided during this pandemic are required. The way they triage, assess and care for patients had to be changed by healthcare systems, using procedures that do not rely on in-person resources. Trends suggested a rising interest in the use of tele-health systems prior to the pandemic. The recent quick policy changes since the COVID-19 diffusion, however, have decreased barriers to access to tele-health and facilitated the use of eHealth as a means of delivering acute, chronic and primary care [80, 81, 82].

Social distancing and quarantining are now common procedures that have been introduced worldwide to limit the people movement with the effect of hindering the COVID-19 spreading. However, the access to services is also reduced. Then, vulnerable individuals whose physiological condition need a regular monitoring, such as frail seniors and chronic patients, require the conventional healthcare model to be changed. A solution consists in adopting digital-health technologies which allow remote health, as well as improve patient outcomes [83]. In this regard, a number of contact-based hospital visits can no longer considered necessary thanks to the recent developments of Internet of Things (IoT) and eHealth technologies [84]. The benefits of employing digital-health solutions consist in facilitating the mitigation healthcare policies during this pandemic by increasing social distancing and by creating a safer environment for health workers and users.

In particular, the adoption of telemedicine for remote monitoring patient during the COVID-19 pandemic has revealed the following strengths [85]:

- it encouraged patients in the conventional healthcare system to be tracked constantly. That also ensures that clinicians would not be disconnected from high-risk communities;
- it avoided the need for non-essential in-person visits, thus decreasing the subjects exposure to COVID-19;
- it reduced the healthcare expenditure, both for users and operators.
- it helped the re-organisation of the clinical staff when provider's resources are highly demanded.
- it provided a critical early warning for COVID-19 symptoms in individuals, especially for asymptomatic or pre-symptomatic individuals.

Beside the great help tele-health is given during the pandemic, the normal management of healthcare systems may also benefits from the adoption of these solutions. A digital-health scenario allows the continuity of care which can avoid negative consequences of late detection and interventions in chronic and routine care. Remote health can offer a more efficient communication channel between providers and medically or socially vulnerable individuals, which often may have a reduced access to health structures. In addition, the maintained patient-clinician relationship enhances prevention of negative conditions, especially in situations where in-person visits are not practical.

Maintaining continuity of care to the extent possible can avoid additional negative consequences from delayed preventive, chronic, or routine care [86].

2.2.3 New Requirements

The World Health Organisation's *Active ageing: a policy framework* document [87] is an international policy tool that has driven progress on ageing since 2002. The principle of active ageing arose in an effort to coherently tie together strongly compartmentalised policy spheres. Active ageing was described in this context as the process of improving well-being, engagement and protection opportunities to increase the quality of life at older age. The importance of a multi-sector action to ensure that older individuals remain a link for their families, communities and economies are particularly stressed. The action should involve six main determinants: economy, behaviour, personal and social life, health and social services, and the physical environment. The recommended components for a health policy response are also defined under the WHO policy framework:

- prevent and reduce the burden of excess disabilities, chronic disease and premature mortality;
- reduce risk factors associated with major diseases and increase factors that protect health throughout the life course;

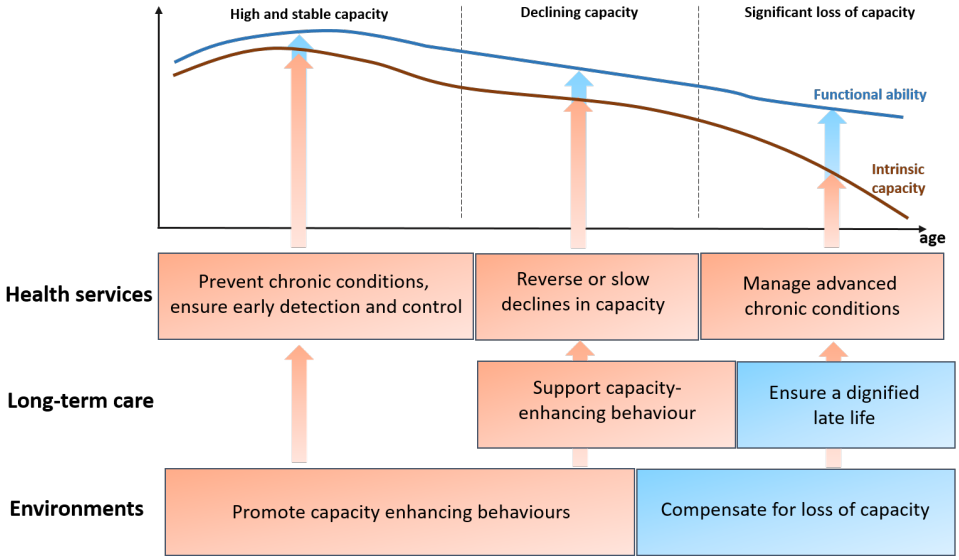

Figure 2.6: Opportunities for public-health action across the life course, adapted from https://www.who.int/ageing/publications/world-report-2015/en/.

- develop a continuum of affordable, accessible, high-quality and age-friendly health and social services that address the needs and rights of people as they age;
- provide training and education to caregivers.

Prevention and early detection of negative conditions is critical for health systems to deal with the upcoming huge demand for long-term care services. To reduce the risk factors, environmental interventions, both by developing personal expertise and awareness, would be key in fostering healthier lifestyles, for example by introducing wider environmental strategies, such as taxing cigarettes or creating clean and friendly environments for physical exercise. The public health framework, proposed in [1], is represented in Fig. 2.6. In this groundwork, a new focus is given to the public-health interventions addressing the part of the population with declining capacities. At this stage, illnesses may have been identified, and health systems' effort will normally move from prevention or cure to mitigating the effects of these disorders on the overall ability of a person. The main objectives therefore will be that to slow or maybe reverse the capacity degradation. In addition, as capability decreases, the role of the environment in facilitating functional ability will increase, with interventions that help individuals resolve the gradually rising problems. The task of long-term care services is to allow the elderly to retain a degree of functional ability that is compliant with their civil rights, opportunities and dignity.

Conventional Care	**Older-person-centred and integrated care**
Focuses on a health condition/s	Focuses on people and their goals
Goal is disease management or cure	Goal is maximising intrinsic capacity
Older person is regarded as a passive recipient of care	Older person is an active participant in care planning and self-management
Care is fragmented across conditions, health workers, settings and life course	Care is integrated across conditions, health workers, settings and life course
Links with health care and long-term care are limited or non-existent	Links with health care and long-term care exist and are strong
Ageing is considered to be a pathological state	Ageing is considered to be a normal and valued part of the life course

Table 2.1: Differences between conventional care versus older-centred and integrated care, adapted from https://www.who.int/ageing/publications/world-report-2015/en/.

The accomplishment of the targets outlined in the WHO's *Active ageing* [87] is not just a matter of improving what has already been achieved. It requires structural reform because, currently, health services in high-income countries are much more designed to treat acute diseases than to control and mitigate the effects of chronic diseases, that are common in older people [88]. The WHO report [1] in 2015, focuses the attention to the central issue of approaching public care from the viewpoint of the health dynamics of an elder individual rather than from the illness or co-morbidity they can experience at a particular point in time. Evidence suggests that the safest approach to apply this complex continuum of treatments across the elder's lives is a person-centred and integrated care [89, 90]. At a clinical level, integrated care relates to a strict cooperation and embedding among different domain areas and sites that are needed to provide long-term support, in addition to administering programs according to the needs of persons during the life cycle [91]. The WHO suggested the key point to focus the attention in reformulating the healthcare model towards a more ageing-aligned system. A synthetic comparison of the characteristics between the conventional care and older-person-centred care have been proposed by the WHO in [1] and they are reported in Tab. 2.1.

2.2.4 Ageing in Place

The living environment usually undergoes modifications to adapt to functional changes caused by the decline in capacity of the ageing individual it hosts, but more often decline require elders to move towards a more supportive accommodation [92]. Older people tend to see their current home or neighbourhood as providing the benefits of retaining a perception of stability and the feeling of being connected to their sense of identification and autonomy [93]. However, conventional institutions are perceived as an obstacle for the social integration of elders with declined functional abilities [94]. Ageing in place is a policy solution to ageing that seeks to strengthen the capacity of older persons to live in their homes and neighbourhoods comfortably and independently. It is also seen as safer for aged people and may also provide major budgetary gains in terms of spending on health services [95].

Ageing in place is strongly supported by growing technologies which allow communication, monitoring and remote assistance. In the next future, this solution will be gradually more relevant and accessible for people. However, as other delicate ageing policies, ageing in place should not be considered as the ultimate solution. However it must be said that in some situations, for isolated elderly persons, for those living in inadequate accommodation or for those living in less comfortable areas, for example, it may not be the primary objective [96].

2.3 Home Monitoring and Early Detection of Decline

The quality of life during ageing is closely related to the chance of receiving an adequate intervention as soon as potentials conditions might arise. Actually, much depends on the possibility to early detect diseases, by the subject itself, by the health services or by the living environment [62]. However the decline trajectories with age can show a variety of different behaviours and there are many existing factors which may have an influence. In general, the intrinsic capacity is likely to slightly diverge in the pre-clinical stage of a chronic disorder, while the differences start to became more evident as conditions progress [18]. Solutions which increases the possibility of spotting the early signs of decline have to be encouraged. Monitoring of daily life activities, for example, might give a closer look to the elder's health status and, together with remote health technologies, it promotes ageing in place. The next sections will treat these topics and will supply guidelines for an effective design of home monitoring systems.

2.3.1 Monitoring of Daily-Life Activities

A large part of the information about the degree of the intrinsic capacity comes from the study of the life-span where major declines in functioning are encountered, often by measuring the activities of daily living (ADLs) or instrumental activities of daily living (IADLs), when the manipulation of common tools is involved. Activity monitoring is usually employed to asses performance in the real world and gain quantitative knowledge about common daily activities [97]. The performance assessment through ADL and IADL in the real world implies the assumption of some considerations which are related to the environment it is applied. It is generally operated in a more or less unconstrained setting that might influence and move the observable measures from the standard references [1].

Research in this field is underway [98] and, while no widely recognised techniques for remotely assessing intrinsic capacity are available, some approaches designed to approximate broad aspects of disability offer a valuable starting point for further study. During 2002-2004, the WHO performed a survey, *World Health Survey* [99], in each country to estimate the status of health, using a series of questions spanning eight areas that could relate both to the intrinsic capacity and functional ability. Impairments in performing basic activities, such as working and doing house-holding activities, moving, keeping personal appearance, concentrating, learning, social involvement and responding to adverse situation, were measured and used to estimate the individual's health status as the average of these items in the life-span. The average health scores and their variation across high-, middle- and low-income countries are represented in Fig. 2.7. The study of the people daily-life activities indicates that, except in low-income countries, the health score remains relatively good until the age of 60 years. After that, in all the countries examined, decline starts to grow more consistently.

2.3.2 Supporting Ageing in Place

Most of the seniors prefers to spend the following late years of their life in their own homes, or at least within their own community [100]. It is therefore important to guarantee protection and the required interventions, even at distance. Information and computer technology in health care, eHealth, is key to enable health systems both to shift towards a more integrated and person-centred care and to remotely access patients health information [101]. The advantages that eHealth and telemedicine provide in healthcare has been confirmed in Europe to enhance the efficiency by 20% [101]. Electronic health records and associated health information systems help to better point the needs of elderly patients, plan their care over time, follow different approaches to therapy and analyse patient outcomes. Collabora-

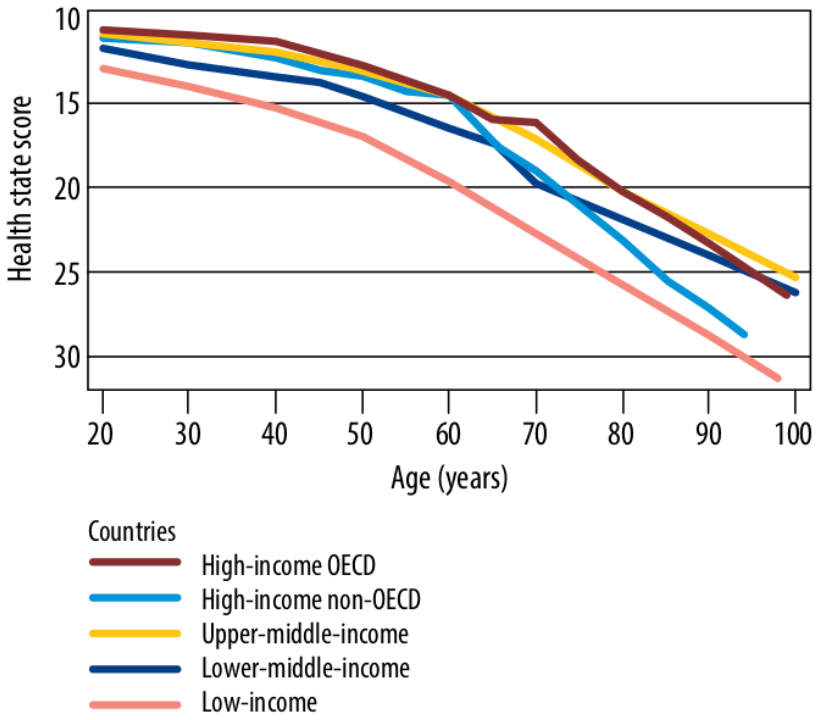

Figure 2.7: Health state score by age in OECD (Organisation for Economic Co-operation and Development) and non-OECD countries, from http://www.who.int/healthinfo/survey/en/, accessed 23 June 2015.

tion between health professionals and between health care teams and their customers, who may be based in different environments or geographical areas, may even be encouraged. Wearable devices can also be used to collect health and physical activity data, such as bio-metric quantities, bio-signals, sleeping activity and gait.

For a consistent and reliable use of these devices however, standard metrics need to be generally accepted and widely used. Some studied also investigated the validity of self-administered digital cognitive tests implemented on mobile devices, such as smartphones an tablets [25]. Results of these tests, performed by the elders in their home, would be sent to clinicians, via telematic forwarding, and a post-hoc analysis could help the early diagnosis of mental disorders without the need for the elderly to often reach the clinical facilities for the clinical evaluations.

2.3.3 Home-Based Health Monitoring Systems Design

Information and communication technology today offer a wide range of solutions that can help enabling remote health monitoring and daily activity tracking. Internet of things (IoT) [102] play a major role in defining those so

called pervasive technologies, which exploit the synergy between sensors and networks to remotely monitor large number of variables in different contexts. A remote and multifaceted monitoring system allows the immediate forwarding of alarms that trigger interventions where necessary or drive automatic regulation mechanisms. In the domestic environment, *smart home* [103] is a term that refers to the application of IoT in houses to track and set physical quantities such as rooms temperature, humidity, air quality and lights or remotely command connected devices. Smart home technology has been proposed and investigated, in several studies [104], to enable health monitoring and independent-assistive-living for community dwelling elders. More recently, common instrumented devices were included in the home monitoring scenario to integrate the tracking information with ADL and IADL related measurements. For example, in the European project MoveCare [12], sensorized daily-life objects were used to acquire data from their deliberate usage by the elder, such as an electronic ink pen to record handwriting gesture data [105].

In developing eHealth systems, holism is a key concept An holistic approach considers people, technologies an all the system components as interrelated and interconnected, as opposed to a more conventional compartmental vision of the parts [106]. A good fit between human and technological factors optimises the effectiveness of the interventions and better achieves the healthcare goals. To achieve this objective, practical human-centred design methodologies, such as participatory development [107], co-design [108] and persuasive technology [109] are recommended. As claimed by Patrick et al. [110], the drawback of the more standard sequential models in biomedical development is that it reduces the complexity of the problem. It studies the behaviour of the system outside from the actual context and the feasibility is then investigated during a final evaluation stage. The mutual interaction between users, technology and environmental factors, at different levels, can not be captured by a rigid sequential approach [111]. It indeed call for a more flexible and iterative process which should adapt the strategies to the new evidences during development.

The Centre for eHealth Research (CeHRes) Roadmap [112] is a framework which combines these indicated approaches and it is used to plan, coordinate, execute and support the holistic research and development process of eHealth technologies. The Roadmap is construct upon five key notions about the eHealth development:

- it is a participatory process;
- it improves healthcare by creating new infrastructures;
- it closely linked with implementation;
- it coupled with persuasive design;

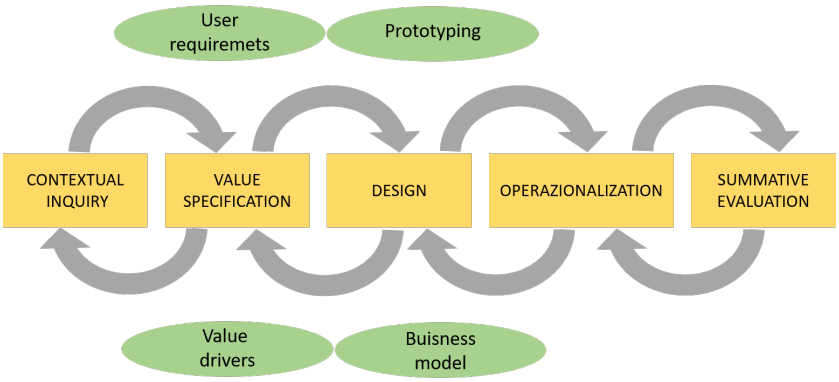

Figure 2.8: A diagram of the Centre for eHealth Research Roadmap, adapted from http//:www.ehealthresearchcenter.org

- it requires iterative evaluation cycles.

The eHealth development, according to the CeHRes Roadmap (Fig. 2.8), is not a sequential process. It is more an iterative chain of development and evaluation phases which ensure that there always be a relation between the next and previous stakeholder perspective, context and outcome.

According to the *Roadmap*, the first step is indicated as *contextual inquiry*. It is dedicated in searching the relevant issues the eHealth solution should address by exploring the context. The technological solution must be shared and by all the stakeholders and it must be suitable for the physical and social environment for which it is intended. This requires therefore the knowledge of users' lifestyle and the behavioural policies of the other stakeholders [113]. Contextual inquiry is the phase in which the team focus on the people and investigate the application context using different tools and methods. First all the stakeholders have to be identified (using for example literature review, expert recommendations or snowball sampling [114]) and then systematic literature review, focus groups, interview and observation of daily practice are carried in order to describe the application domain, with is weak and strong points.

The second step, the *value specification*, elaborates what is needed from the technology. In this phase, the team starts from the issues (or points of improvement) identified before and finds out which are the goals and the specifications the solutions should achieve, according to all the stakeholders. In particular, the value specification focuses on which are the added values a technology might bring to the current setting. The identified values should be then prioritised and a business model of the eHealth solution is drawn up. At last, the values are translated into technological requirements.

The third step is the *design* phase. In order to create solutions that will be used as long as possible, the design phase should be closely linked to the implementation. In fact, issues can remain hidden during design and

only emerge after the technology deployment. Design and implementation coexists through the generation of multiple prototypes, used to iteratively test the results and improve ideas. Prototypes can be also tested with users and other stakeholders to find whether the technology fits their needs and preferences [115]. Another important aspect to consider in the design phase is the technology appeal to the user: the influence of technology on users and their habits must be such as to ensure that it will be readily accepted. This is the goal of the *Persuasive System Design* model which aims at making technology compelling, without being coercive [116]. The persuasive features of a technology enhance its adherence to the context and increase its effects [117]. In the case of eHealth systems, persuasive design can be intended to influence specific behaviours in users (e.g. taking medications, having a healthy nutrition or doing physical activity) with the insertion of behaviour change techniques in the design [106].

Usability tests have to be performed on prototypes to identify defects and collect feedbacks from users. According to a study by Jaspers in 2009 [118], the best outcomes are obtained by performing both expert-based and users-based usability tests. In the first case, expert and designers evaluate the prototypes in relation to the specific project requirements or to task-oriented usability principles (heuristics). In the other case, members of the target group use the prototype in real-world scenarios and the impressions and suggestions of the users are collected.

The fourth phase is named *operationalization* and is the time in which the plan for disseminating and introducing the technology in the practice and in the market is established. A key point to consider is the definition of the technical support strategy for the deployed technology and the possibility to adapt the technology to the changing demands or requirements from stakeholders.

The last phase is the *summative evaluation* where the impact of the technology is determined and the adoption of the system by the users is analysed. For an eHealth systems, the impact can be measured for example by looking at improved users' clinical values, quality of life and adherence to advice or treatments, or at the improved availability, efficiency and interaction of the healthcare provider. The adoption of the technology can be investigated by studying the usability, the frequency and the patterns of use and by counting the number of 'hardcore' and 'dropout' users. In addition, interviews can be used to point out the reason of dropouts and the features that hardcore users appreciated more.

Information on general and outcomes and usability of an eHealth system can be collected with several survey methods: the effect on the health status of patients before and after the use of the system can be evaluated with a *randomised control trials* [119]; usability data and precious insight on the daily use (and impact on user's health-related behaviour) of the system can be obtained by analysing log data and time series collected during an *ecologi-*

cal momentary assessment [120]; and *health technology assessment* [121] can be used to perform a broader evaluation of several issues including social, economical and ethical impacts.

In the specific scenario in which use ADL (or IADL) is included in the home monitoring system, important aspects should be addressed to actually let the operate at their best. The interaction between the users (i.e. older adults), the instruments and the environment need a special focus during the design phase. Yet these aspect represents the major challenge in research for their development.

The main concepts for designing monitoring IoT systems for aged individuals are listed here:

- *Safety and privacy.* Safety and privacy are basic aspects that involve any application of IoT devices in human-living environments [122]. The risks associated with using these tools must be extremely low and the acquired information should not include sensible data as that would represent a barrier to their deployment.

- *Transparency.* Non-invasiveness (or ecology) is a key feature for home-monitoring systems and devices for elders, since the request of elderly users to adopt new behaviours and tasks may make the system feel uncomfortable and intrusive. Especially considering the quite spread technological illiteracy of older adults [123]. The best solution is therefore to adapt the technology to the user's habits, using intelligent methods to obtain the right information about his activity. This branch of research belongs to the wider topic of *ambient intelligence* [124].

- *Data consistency.* The concept of non-invasiveness contrasts with the possibility to apply standard protocols for the remote measurement of biometric quantities or performance in certain tasks. This may implies the absence of reference values for the subject's tracking outcomes in the short and long period. Moreover, the home of the individual represents an uncontrolled environment that imply the presence of a more or less high amount of noise in the measurements over time. Therefore, to ensure a consistent monitoring, analyse longitudinal trends and make further reasoning, the measurements must be reliable and the nominal behaviours, as well as the variability of the metrics should be formerly studied in related control populations.

Chapter 3

Handwriting Analysis in Literature

Ageing in place is a solution to promote healthy ageing in community dwelling elders [95]. From a technological side, eHealth and remote monitoring systems are key to allow a safe life and to ensure adequate health care for older people. Especially for individuals in the age 60-70, for which the state of pre-frailty has a large incidence, a continuous tracking of the health status, by means of IADL monitoring, could be decisive to early diagnose potential age-related diseases [97].

This chapter is divided in two parts. The first, Sec. 3.1, shows the instruments currently employed in the analysis of handwriting and their limitations in the domestic scenario for home monitoring purpose. The second, Sec. 3.2, explains the methods used in literature for the quantitative study of handwriting.

3.1 Limitations in Home Monitoring

Handwriting is an daily-life activity that has been found to undergo changes with advancing age and with the presence of some pathology. Handwriting in fact is a continuous cognitive-motor task acquired during development that requires high skill and cerebral activation [125]. It is a familiar and straightforward activity for almost all literate adults, which was proven to be a very useful biomarker. Indeed, the motor performance required for writing depends upon the coordinated function by the brain, in combination with the neuromuscular and visual systems. This deteriorates, to some extent, in all older adults and even more so when neurological disease is present [8]. For this reason, the analysis of this activity has been leveraged for assessing different conditions. As for neurology, kinematic analysis of handwriting has been used as a clinical tool highly sensitive to even subtle dysfunctions,

Figure 3.1: Examples of commercial devices for writing data acquisition: a digitizing surface, on the left, and a tablet with an electronic pencil, on the right.

particularly useful in Parkinson's disease (PD) [8], Dystonia [13], and Huntington's disease [126] evaluation. Given its fine motor nature, handwriting analysis turned out to be a very useful tool also for the investigation of tremor [127]. In addition, handwriting was studied to discriminate different levels of severity in terms of age-related cognitive decline [10]. According numerous scientific evidences, the characterisation of age-related changes results a potentially useful tool to distinguishing between physiological variations to abnormal changes, possibly related to neurological conditions or cognitive decline.

Although it has been revealed to be a useful indicator of variations linked to ageing or pathological decay, handwriting has not yet been investigated as a home health monitoring tool. The first reason can be ascribed to the lack of an handwriting assessment tool that meets all the aspects of simplicity, non-invasiveness and privacy needed for an elderly home-based monitoring system. Indeed, early work used ink pen and paper notebook to register the subject's writing outcome [8]. On one hand, this approach can be considered worthwhile in the clinical environment, due to its simplicity, since it does not require the support of a technician. On the other hand, the assessment of the paper-and-pen technique requires the expertise of a clinical professional to evaluate the writing outcome without the support of any quantitative data: such an approach does not match the current needs of the health systems that count on the achievements of telemedicine to solve problems such as the limited availability of specialists, the reduced time to conduct such tests, and the difficulty for some patients, especially the older ones, to reach the examination site [128]. For this reason, in most of the recent studies, the paper-and-pen approach was replaced by digitizers and tablets (in Fig. 3.1) able to return the 2D trajectory of the writing trace [10, 129, 130]. The digitisation of data allows extracting quantitative parameters to objectively assess handwriting and achieving remote monitoring of user's performance. However, such an approach is questioned since it constraints

the user to write on a relatively small surface (typically the one of a tablet), that is not the standard writing surface; in this way, the naturalness of the gesture is undermined [131]. As a consequence, this approach lacks of ecological validity, since the experimental context does not match the real-world phenomenon [11]. Moreover, the use of such technology may not be so straightforward, particularly when dealing with elder users, thus requiring the technical support of an operator.

The second reason is that, in an ecological home monitoring setting, users are free to write in different moments, situations and modalities and such uncontrolled context may increase noise and data variability. Indeed, the most totality of the previous research investigated the handwriting assessment in controlled conditions, using standard protocols (such as predefined sentences or patterns to draw[1]). However, in the home-based monitoring, standard protocols should be avoided for the lack of supervision and for the principle of transparency: applying protocols would mean imposing procedures on elderly users that could lower the level of acceptance of the health monitoring system. As a consequence, ecological validity should be ensured by obtaining measurements independent from the modality and the content of the writing activity.

The third reason, the almost absence of golden standards to compare handwriting data collected in a domestic and uncontrolled environment suggests the use of unsupervised or semi-supervised approaches to capture any non-physiological sign of decline in elders. Anomaly detection (AD) techniques [132] are used to identify irregularities from data, even when the knowledge about them is partial or missing. Anomalies observed in the user's writing data could indicate a deviation from a normal ageing process.

3.2 Quantitative Analysis of Handwriting

The analysis of handwriting consists in finding quantitative descriptors of the writing activity, in the written trace or in the writer's hand gesture. These quantities are treated as indicators which characterise the subjects' personal handwriting. Since handwriting is a complex learned skill, its characteristics tend to change when the writer's motor control system undergoes modifications, for example during the learning process, or when experiencing age-related decline. Changes at higher or lower level in the motor control system implicate different effects in the writing activity.

Ageing and age-related pathology are associated with a decline in the individual's cognitive and motor systems [133, 134]. A large body of research investigated how handwriting features are altered during the healthy ageing

[1] See Section 3.2

process and in presence of a pathological age-decline [135, 136]. Most of the handwriting features treated in literature can be grouped into four domains:

- time-domain,
- kinematic/dynamic-domain,
- complex-domain.
- frequency-domain ,

The *time-domain* features include those indicators that describe temporal aspects of the handwriting activity. It has been shown that the time taken to write the strokes that make up words and letters (*on-paper time*), the time between the execution of two consecutive strokes or between different words (*in-air time*), and the relationship between these two quantities (*in-air/on-paper ratio*) may reveal decline. In 2012, Rosenblum et al. [129] studied the changes in Executive Functions and in handwriting related to ageing. They included 80 healthy participants, aged from 31 to 76+, performing a handwriting task on a digitizer that acquired the kinematics of the writing trace. They showed that the variance of temporal characteristics of handwriting, such as *on-paper time*, *in-air time*, and the *in-air/on-paper ratio* were predicted by age. In particular, age predicted the 32% of the variance of the *in-air time*. Especially for those over 60s, higher value of *in-air time* and *in-air/on-paper time* were found. Authors assumed that an increase in writing in-air time can suggest a decline in the fluency and efficacy of the production of handwriting ascribed to age. In addition, *in-air time* may not only be an indicator of functional deterioration, but may more precisely be a sign of problems in the planning of spatial and temporal activities. The *in-air/on-paper ratio* outcomes uphold the idea that with increasing age, more time is needed to plan the next steps in the executions of handwriting. Another study from the same group in 2010, showed that *in-air time* in handwriting significantly increased in major depressive disorder patients, with respect to control subjects [10]. In the same research, temporal and spatial measurements, together with writing pressure acquired when participants were asked to write their name showed an accuracy score of 82% in the classification of the participants.

As already seen in the time-domain features of handwriting, the fluency of the writing activity is expected to decrease with age or, more in general, when functional decline takes place. A decreased fluency in handwriting results in a reduced regularity which can be further investigated in the *kinematic- and dynamic-domain*. The number of changes in velocity and acceleration in handwriting were studied by Tucha et al. in 2012 [137]. The decrease in handwriting fluency was investigated in individuals with lesion of the precentral region of the Left Hemisphere by measuring the number

of changes in velocity and accelerations during writing. The values resulted significantly higher in patients than in healthy subjects, suggesting that impairments increase irregularity. In 1997, Walton [8] studied the changes in the writing force and in the number of changes in the writing force due to ageing and PD patients. He showed that force tend to be weaker in presence of functional decline and it also tends to became more uniform, by reducing its variations within strokes.

The *complex-domain* indicators describe the non-linear features of the handwriting activity. Generally, the complex-domain indicator evaluate the presence of the tremor components in the handwriting signals. When tremor is present, oscillatory behaviours superimpose to the writing dynamics and it results in an increase predictability of future trajectories. Hong et al. [15] assumed that a more complex behaviour is visible in healthy-young writer and such complexity is expected to decrease along the decline process. Longstaff and Heath showed that handwriting could manifest complex dynamics [138]. They studied the non-linear features of the trace velocity of digitally recorded handwritten words, performed by healthy subjects. They obtained relatively small fractal correlation dimensions (with average value of 3.24 for the horizontal speed and 2.9 for the vertical speed) and positive maximum Lyapunov exponent (0.114 and 0.118) from the written trace velocity. Their results showed that handwriting in healthy subjects can show the behaviour of a chaotic system of low dimensionality and they hypothesised that complexity can decrease with functional decline. Meigal et al. used non-linear quantities to describe the behaviour expressed by PD patients and control subjects in hand movements [139]. Their findings reinforced the hypothesis of a reduced complexity associated with a reduction in the individual capacity, as they discover that the value of the of determinism and recurrence ratio, calculated from the acceleration signals, were significantly ($p<0.001$ with significance defined at the 0.05 level) higher in presence of Parkinson's disease, and their trend was increasing with age. In fact, determinism measures the predictability of a time series, while the recurrence ratio quantifies the tendency of a signal to express similar patterns.

The *frequency-domain* indicators refers to the tremor factor in handwriting. Alty et al., in 2017 [127], published a comprehensive review on the tremor features of four tremor diseases (essential tremor (ET), dystonic tremor (DT), Parkinson's disease (PD) and functional tremor (FT)) in pen and paper tasks. They qualitatively analysed the written trace and the drawings on paper and they characterised the tremor types for each of the four conditions. Their findings are reported in Tab. 3.1 for the handwriting task. The table includes tremor features, but also letter size, pen pressure and the tendency to deteriorate within the disease course. Beside a more qualitative description of tremor, quantitative hand tremor indicators can be found in

Handwriting	**ET**	**DT**	**PD**	**FT**
Size	Normal or large	normal	small	variable
Tremor	Regular amplitude and frequency	Irregular jerky amplitude and frequency	Regular frequency amplitude	Variable
Tremor in letter sections	Vertical strokes; unidirectional axis	All sections of letters; multidirectional axis	Vertical strokes; unidirectional axis	Variable
Progressive decline	No	Yes	Sometimes	Not usually
Pen pressure	Normal	Hard	Normal	Normal

Table 3.1: Characteristics of tremor types seen in writing, adapted from Alty et al. 2017.

the study by Hong et al., in 2007, [15]. Authors tested the hypothesis that ageing reduces the coupling between motor system compartments, thus causing a reduction of behaviour complexity. They measured this complexity by calculating the approximate entropy of finger acceleration during rhythmical movement and postural tasks in subjects divided into three different age groups: 18-23, 60-65 and 70-75 years old. Entropy is related to the number of possible behaviours that an individual, in a stable physical-cognitive condition, can express during the execution of a functional task. This is also related to its ability to adapt and react in different situations. Whit age, the range of behaviours tend to decrease and the rising oscillatory tremor components enhance the regularity of motion. Results confirmed that the regularity of the acceleration dynamics increased during postural tremor and movement in the mid-old and older subjects.

Other tremor indicators can be found in the study by Edwards and Beuter [140]. They measured the central tendency of tremor, as the median frequency, in postural tasks using using a laser system as well as a portable commercial tremor analysis system in healthy individuals and Parkinson's disease patients. They showed that it is generally possible to locate tremor components generated by particular dynamic conditions of the gesture in specific frequency bands. In particular, the range between 3-4 Hz have been found in individuals with cerebellar deficits and the parkinsonian tremors are considered to be in the 4-6 Hz.

A more recent study by Garre-Olmo et al., in 2017 [141], combined the use of temporal, kinematic/Dynamic and complex indicators to discriminate between AD and mild cognitive impaired (MCI) patients from control subjects. They involved 52 participants (23 AD, 12 MCI, and 17 healthy controls) with a mean age of 69.7 years who were asked to perform three

Handwriting task	**Healthy vs MCI vs AD**	**Healthy vs impaired (MCI+AD)**	**Healthy vs MCI**	**MCI vs AD**
Spontaneous	age complexity	speed complexity	age pressure acceleration complexity	age pressure complexity
Copied	age pressure complexity	complexity	pressure complexity	age pressure acceleration speed on-air time
Dictated	complexity	pressure speed acceleration complexity	pressure acceleration speed complexity	age pressure acceleration speed

Table 3.2: Discriminant handwriting indicators for healthy, MCI and AD patients for three handwriting tasks, adapted from Garre-Olmo et al. 2017.

writing tasks (copy one sentence, write a dictated sentence and an own sentence) and other drawings exercises. Data were recorded using an electronic digitizer system. They used discriminant analysis to explore the value of handwriting indicator, including writing pressure, acceleration, speed, complexity, in-air time and tilt angle, to classify participants depending on their degree of cognitive impairment. They showed that the degree of correct classification was task and group dependent and it ranged from 63.5% and 100% of accuracy. Table 3.2 report the most sensible indicators in the classification for task and groups involved, The kinematic indicators showed higher specificity in discriminating between normal and impaired condition. The highest sensitivity was obtained in distinguishing the level of severity (MCI from AD patients).

A synoptic table of the most relevant indicators treated in the literature is shown in Tab. 3.3. For each item, the domain an the tendency to increase/decrease with age or pathological decline is indicated.

The quantitative analysis of handwriting has been treated in several contexts in literature, most of which implied the use of specific protocols in controlled settings. Therefore, the validity of results can not be generalised for all the situations. In uncontrolled environments, as in the case of remote home-based handwriting assessment, three specifications should be investi-

Indicator	**Domain**	**Trend with age and/or pathology**	**Reference**
in-air time	time	increase	Rosenblum et al. 2010/12
on-paper time	time	-	Rosenblum et al. 2012
in-air/on-paper time	time	decrease	Rosenblum et al. 2012
modal frequency	frequency	-	Edwards and Beuter, 1999
approximate entropy	frequency/ complex	decrease	Honget al., 2007
writing force	kinematics/ dynamics	decrease	Walton, 1997
# changes in force	kinematics/ dynamics	decrease	Walton, 1997
# changes in velocity	kinematics/ dynamics	increase	Tucha et al. in 2012
# changes in acceleration	kinematics/ dynamics	increase	Tucha et al. in 2012
determinism	complex	increase	Meigal et al. 2012
recurrence ratio	complex	increase	Meigal et al. 2012
tilt angle	kinematics/ dynamics	-	Garre-Olmo et al. 2017

Table 3.3: Main handwriting indicators in literature.

gated and satisfied before validating an handwriting indicator:

- the indicator must describe an invariant characteristic of handwriting in an individual in a stable health situation [2];
- the indicator behaviour must be independent w.r.t. the handwriting modality (i.e. right or left hand, pen grip) and the writing content;
- the indicator must be sensible to changes in the individual's handwriting due to ageing or pathological degradation.

This work proposes a novel technological solution expressly designed for the ecological monitoring of handwriting in community-dwelling older adults: a novel smart ink pen that allows users to write on a common piece of paper while acquiring motion and force data. The writing gesture dynamics recording instrument is designed to be easy to use as it is able to autonomously manage data collection and wireless data transmission. From the user's point of view, the device almost appears like a common pen with a replaceable ink refill and a rechargeable integrated battery, to encourage its use during normal daily activities. The signals acquired with this tool allow the calculation

[2]Stable is referred to a reasonable limited amount of time

of all the quantities considered in the literature as useful indicators for tracking the subject's intrinsic capacity.

Once verified their reliability in the uncontrolled environment, the handwriting indicators are studied as biomarkers for age- and pathological-related writing characteristics.The handwriting indicators of a pen user can be monitored over time to identify anomalies which may suggest an abnormal decline in capacity.

Chapter 4

The Smart Ink-Pen

To realise a smart writing object that resembles a common ink pen, a careful design of the micro-architecture to sense and transmit the data has been carried out[1]. The developed pen has the following dimensions: height 147 mm, maximum diameter 14.65 mm and weight 48 g. An image of the pen and its internal components is reported in Fig. 4.1.

In this chapter the hardware and the software architecture of the pen are described in Sections 4.1 and 4.2 respectively; Sec. 4.3 shows the procedures performed to test and validate the instrument; Sec. 4.4 reports the handwriting indicators statistics and reliability study; and sec Sec. 4.5 shows the on-filed investigation of the smart pen usability.

4.1 Hardware

The pen incorporates the internal electronic components to acquire, store, and send the handwriting data. The electronics is integrated in three Printed Circuit Boards (PCB) located in the upper part of the pen. PCB1 has a dimension of 28.5 x 8 mm and is the core PCB; it includes an ultra-low power Cortex^{TM} 32bit CPU (STM32L476) to acquire, filter, and transmit the signal, a BlueNRG-MS single-mode network processor (compliant with Bluetooth specification v4.1) to implement BLE connectivity, and a 1MB flash memory for storage purposes. In terms of sensors, inertial signals are acquired through 3D linear accelerometers and gyroscopes (LSM6DSM iNEMO® 6DoF), while the writing force exerted on the tip is measured through a miniaturised load cell (FC8E by Forsentek; ∅ 1.6 mm; 50 N capacity) mounted with the lower face pressing on the refill stopper. The PCB2 includes a rechargeable Li-Ion PIN-Type battery by Panasonic, with a coaxial power connector accessible from the pen cap, and the battery protection circuit. PCB3, with a rounded shape, is placed in the load cell holder and

[1] A patent regarding the smart ink pen has been submitted [142]

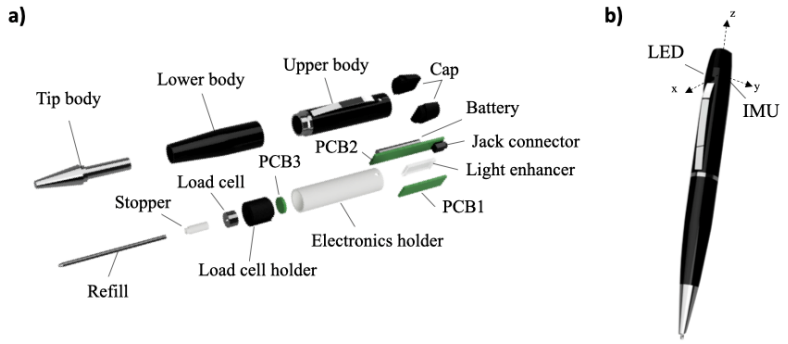

Figure 4.1: (a) A rendering image of the smart ink pen and its internal components. (b) An external view of the smart ink pen.

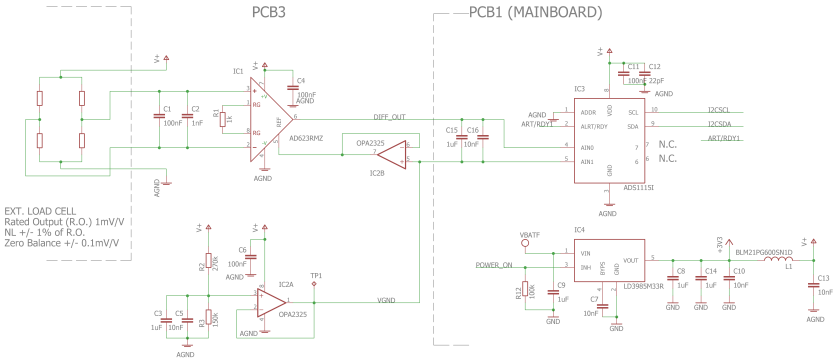

Figure 4.2: Conditioning circuits of the smart pen.

includes the pre-amplification circuit for the load cell data. Pen functioning in self-operated through movement detection or BLE connection request, therefore no activation button is available for the end-user. Nonetheless, a LED is visible to indicate the operating mode and the battery state of the pen. All the electronics is protected and securely fixed inside a 3D printed plastic case.

Figure 4.2 shows the signal conditioning circuit for the load cell. The circuit comprises a low noise instrumentation amplifier (AD623RMZ (IC1) hosted by PCB3), connected to a 16bit high-quality, low noise, low power fully differential ADC (ADS1115I, (IC3) hosted by PCB1), both powered by a 3.3VDC single supply low dropout voltage regulator (LD3985M33R, (IC4) hosted by PCB1). The IC4 receives energy directly from the battery (PCB2), thus reducing the possible interference arising from digital and wireless parts of the mainboard (PCB1).

The instrumentation amplifier (IC1) has a gain set at 101, and a low noise, low power RRIO dual operational amplifier. The differential input filter has been set to 1.5kHz, to allow a first stage filtering before the con-

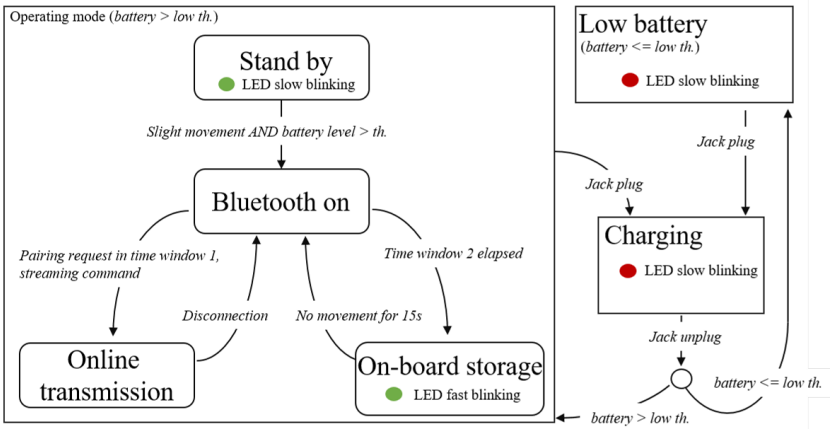

Figure 4.3: Block diagram of the firmware operating modes.

nection wires to the IC1. The output baseline (Virtual Ground VGND) has been set to VCC/3 by the IC2A, to allow a fully differential analog to digital conversion without losing the possibility to investigate possible pre-loading effects of the load cell caused by mechanic assembly. The output of IC1 and the VGND signals coming from IC2 have been connected to the main-board (PCB1) respectively to the positive (AIN1+) and negative (AIN1-) channel inputs and of the ADC (IC3). The ADC samples at 50sps. The internal PGA of the ADC has been programmed to 2, providing a total gain of 13,3mV/N. An anti-aliasing filter has been placed near 0.1Hz to preserve the signal information.

4.2 Firmware

The firmware was designed with the aim of maximising battery duration and usability of the smart pen for not expert users, such as older adults. To extend the possible applications of the pen, two modalities of data transfer are envisaged: i) online data transmission: the micro-controller reads the data from the sensors and transmits them in real-time through BLE; ii) on-board storage: when the object is used, sensors data are sampled and saved on the flash memory, and downloaded afterward.

The firmware comprises six operating modes and Fig. 4.3 shows its block diagram representation. When in the *Stand-by* state, the battery is set to saving mode and, just to advice that the pen is available, a green LED blinks at a low frequency. Only when a slight movement of the object is detected by the accelerometer, and the charge level is over a specific threshold (to prevent using the object when the battery is too low), the BLE module of the object is switched on. In this *Bluetooth ON* state, the pen is available for any pairing connection within a certain time window. If a pairing request is received

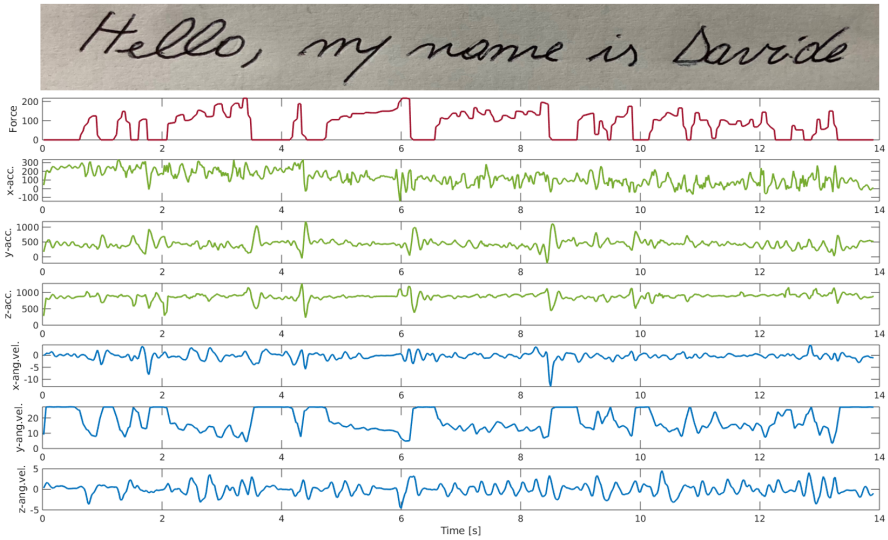

Figure 4.4: The signal acquired using the pen during handwriting. The upper panel shows the written trace on paper, other panels report the signals collected using on-line storage mode of the smart pen: the writing force on the pen tip (non-scaled units), three-axis accelerations [m/s^2] and three-axis angular velocities [rad/s].

within the time window, followed by a start streaming command, the *Online Transmission* state is enabled, and the object transmits the information packages at a frequency of 50 Hz. The object returns to the *Bluetooth ON* state once a disconnection request is received. Otherwise, if (in the *Bluetooth ON* state) a pairing connection is not received within the time window and the object is moved by the user, the transition to the *On-board Storage* state is triggered. At this point, the object starts storing data packages on-board at 50 Hz, and stops when left stationary for 15 seconds thus triggering a transaction to the *Bluetooth ON* state. When the battery level is below a certain threshold, the object goes in the *Low Battery* state in which a red LED blinks to notify the user that charge is needed. Every time the pen is connected to a charger, the state moves to *Charging*, with the red LED blinking at a higher frequency.

The data package, sampled at 50 Hz, includes a timestamp, the 3-axis acceleration [m/s^2], the 3-axis angular velocity [rad/sec], and the writing force signal exerted on the pen tip. Finally, a software was created to automate the data download and processing procedures. An graphic example of the signal acquired during an handwriting activity using the pen is shown in Fig. 4.4.

4.3 Pen Testing and Validation

This section includes the apparatus, the protocols, and the analyses aimed at validating the pen sensors and algorithms.

4.3.1 Tip Force Static Calibration

Methods

This test has the twofold aim of verifying the linearity of the writing force measurements in the range of the force values exerted during handwriting, and estimating the optimal calibration parameters for the conversion of the pen tip force signal to Newton (N) units. The set-up shown in Fig. 4.5 was devised to keep the pen in vertical position and to place the test weights at the top; it includes two 3D printed arms to hold the device with the minimum friction, and a circular flat base to accomodate the weights. From the pen load cell, the normal component of the force signal, F, applied to the tip was acquired while increasing the testing weight from 0g to 50g, with steps of 10g, and from 50g to 500g, with steps of 50g. For each weight increase, the related F measure was obtained as the mean value of the signal over 50 samples after the transient phase. The linear regression was computed between the measurements of the pen force F and the corresponding testing weights placed on the top of the pen; the selected loss function was the Mean Squared Error (MSE).

Results

A strongly significant positive linear regression, showed in Fig. 4.6, was observed between the test weights, in grams (g), and the pen measurements (non-scaled units) with a R_2 score of 0.99. The linear coefficients, m (slope) and b (intercept), for the force signal calibration were equal to 0.62 and -369.7g, respectively.

4.3.2 Tip Force Dynamic Calibration

Methods

The writing force signal was validated in dynamic conditions (i.e. during the handwriting activity), comparing the pen force signal F with an external force reference signal F_{ext}, obtained through the M01-500N Golden Type (by Sunrise Instruments©) load cell sensor. The protocol consisted of a 2-minute free writing test performed by a healthy adult, who used the smart pen to write over the load cell surface, as shown in Fig. 4.7. The two signals were synchronised using the peak of cross-correlation. The Pearson's correlation

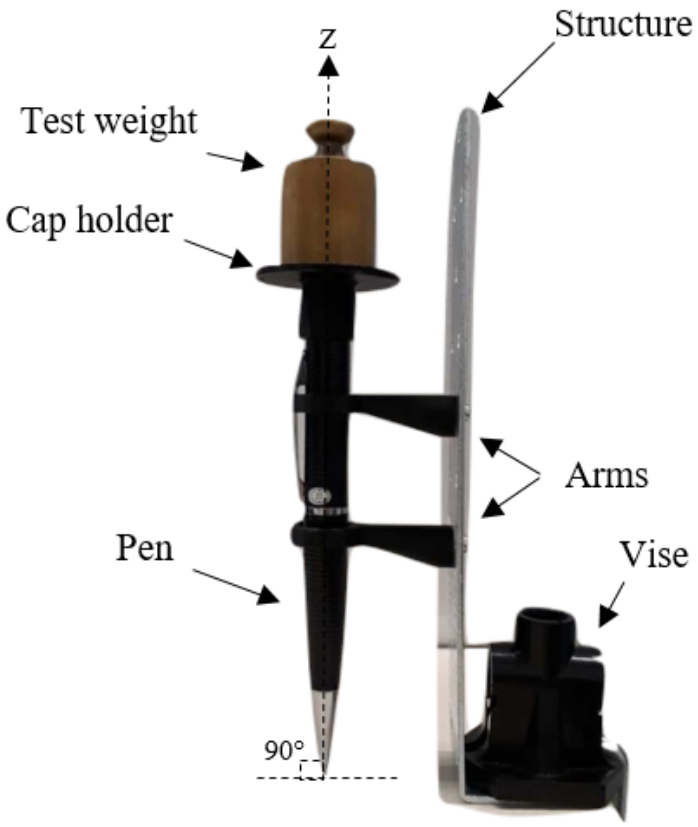

Figure 4.5: Experimental set-up for the pen static tip force calibration.

coefficient ρ_d was evaluated between F and F_{ext} during the writing task.

Results

The correlation ρ_d between F and F ext was significant and equal to 0.96. Figure 4.8 shows the pen force signal and the external sensor force signal superimposed acquired in the first 30 seconds of the writing task. Both signals are normalised with respect to their maximum value to better visualise the comparison. As noticeable, the writing force signal F recorded with the pen accurately matched the reference signal F_{ext} from the load cell, with a weak smoothing effect on the fastest components.

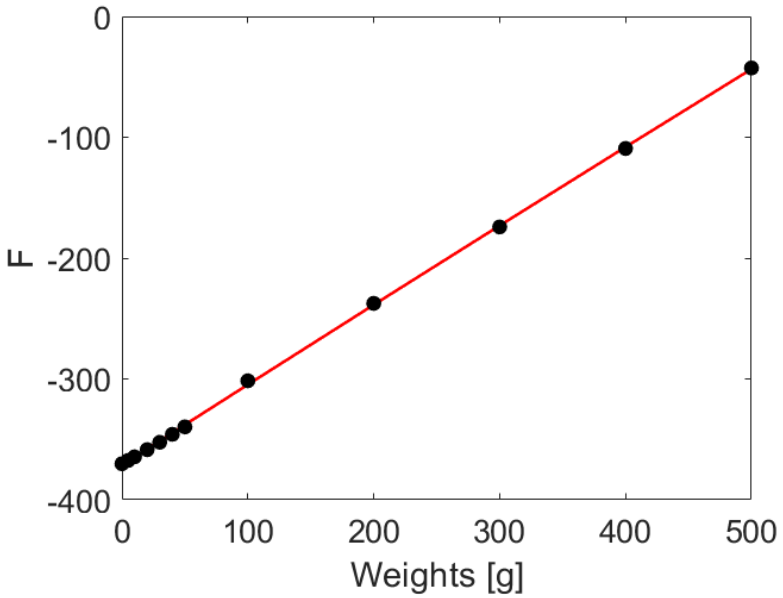

Figure 4.6: Linear regression between the test weights (x-axis) and the pen measurements (y-axis).

4.3.3 Tilt Angle Validation

Methods

The tilt angle θ_{pen} obtained from the pen's motion sensors was compared with a reference tilt angle θ_{ext} obtained from the optoelectronic motion capture system SMART DX 400 (BTS SPA, IT). To this aim, a 3D-printed tool with 3 retroreflective markers was built and positioned on the top of the pen to record the orientation of the pen body, as shown in Fig. 4.9. The acquisition protocol included two different trials performed by a healthy adult: in the first one, the subject was requested to draw six straight lines on a sheet of paper, three with the smart pen tilt kept approximately at 45°, and three at 70°. In the second trial, the subject was asked to write freely for three minutes, and no constraint on the pen tilt was imposed.

In quasi-static conditions (i.e. when the pen is moved very slowly), the tilt angle, θ_{pen}, was computed using the approximation in Eq 4.1:

$$\theta_{acc} = sin^{-1}\left(\frac{a_z}{g}\right), \tag{4.1}$$

where a z was the z-axis acceleration signal and g=9.81 m/s was the gravitational acceleration; for small tilt angles ($\theta << 1°$), the relation was linearized as in Eq. 4.2:

$$\theta_{acc} \approx \frac{a_z}{g}. \tag{4.2}$$

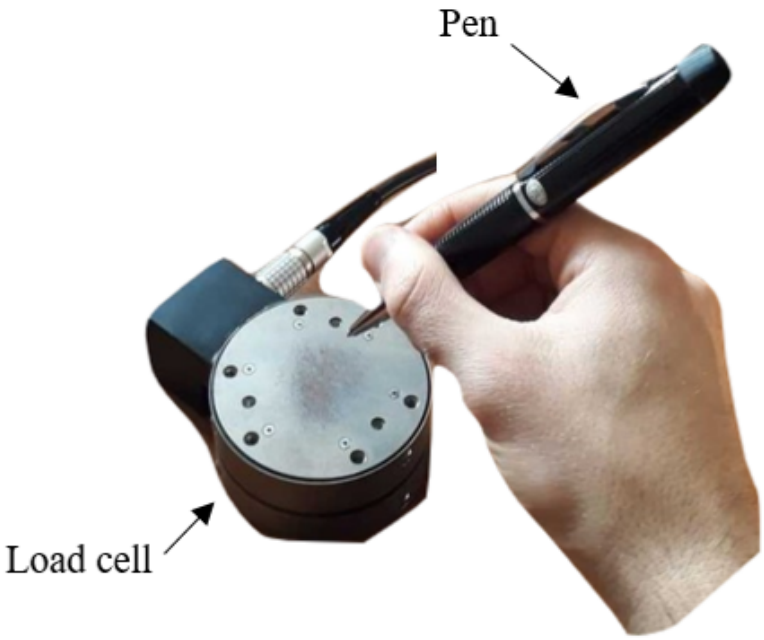

Figure 4.7: Experimental set-up for the pen static tip force calibration.

The estimate tilt angle pen, θ_{pen}, corresponds to the approximation, θ_{acc}, for slow movements.

In non-stationary conditions, instead, a low-pass filtered (cut off 10 Hz) a_z was considered in the computation of θ_{acc} to get rid of the z-axis acceleration components not related the gravity load. Indeed, at high frequencies, the variation of the tilt angle is coupled with non-negligible accelerations which lead to consistent errors in the estimation of the tilt angle from the accelerometers [143]. Furthermore, to avoid the estimate to diverge, the estimate from the accelerometer signal of the tilt angle was combined with its rate estimate from the gyroscopes signals, $\dot{\theta}_{gyr}$, obtained by integrating (in time domain) the angular velocity in the x- and y-axis. The sensor fusion filter in Eq. 4.3, [143] was used to obtain the estimate of the tilt angle rate $\dot{\theta}_{gyr}$:

$$\dot{\theta}_{pen} = k_1\dot{\theta}_{gyr} + k_2(\theta_{acc} - \theta_{pen}), \tag{4.3}$$

where, k_1 and k_2 were two constants empirically set to 1.5 and 0.4, respec-

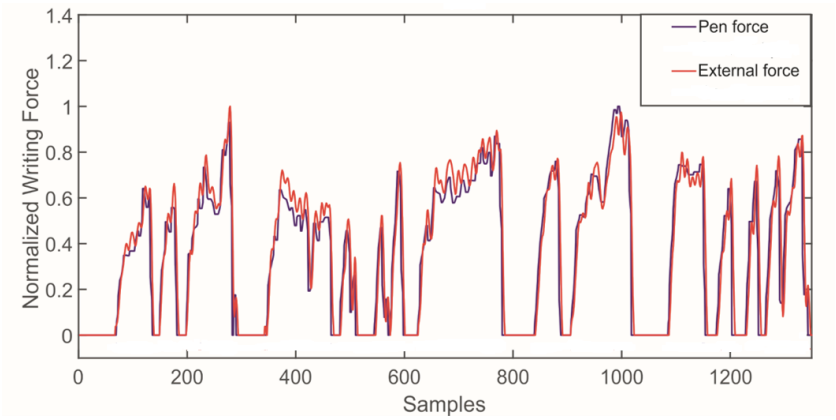

Figure 4.8: Aligned and overlapped writing force signals acquired with the pen (blue) and with the external load cell used as reference (red).

tively, to achieve the minimum discrepancy between the estimate and the measured tilt angle.

Finally, two tilt angle signals were compared by calculating the Pearson's correlation coefficient ρ_t and the Root Mean Squared Error (RMSE) between θ_{pen} and θ_{ext}, during the writing tasks.

Results

The results of the tilt angle validation are shown in Fig. 4.10. The results in the first writing task reported a significant correlation coefficient ρ_t of 0.89 and a RMSE value of 6.3° between the estimate tilt angle from the pen and the reference angle obtained with the optoelectronic system. In the second task, correlation was significant with ρ_t equal to 0.78; a RMSE of only 3.8° was reported.

4.3.4 Segmentation Into Strokes

Methods

The algorithm for the segmentation of the signals into strokes is the preliminary step for the calculation of the handwriting indicators; a stroke was defined as an interval in which the pen writing force was non-zero. For its validation, the segmentation into strokes computed from the pen force signal, F, was compared with the one extracted from the external load cell reference signal. The signals were acquired using the experimental setup showed in Fig. 4.8 and repeated ten times with different smart pen prototypes. Ten different pen prototypes were used to validate the robustness of the segmentation algorithm considering possible differences due to device construction

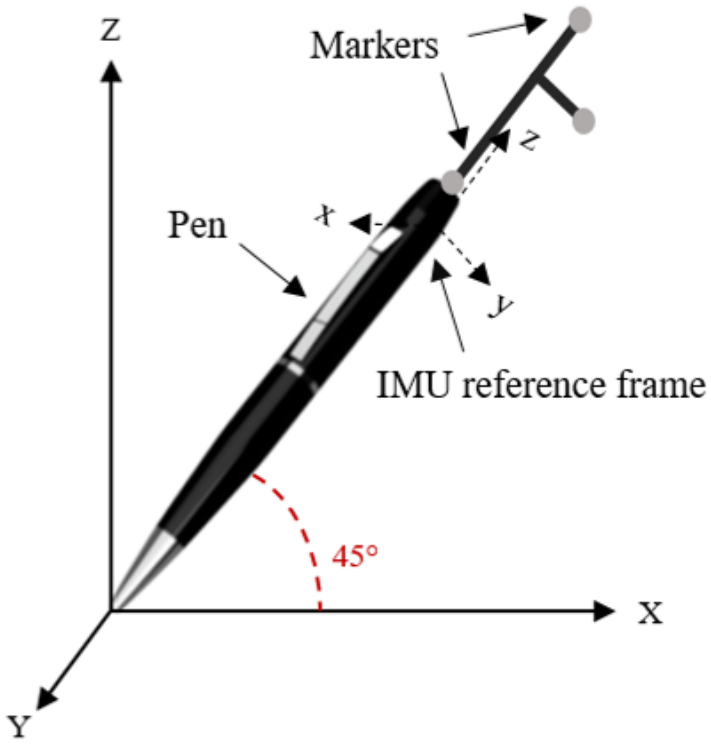

Figure 4.9: Experimental set-up for the tilt angle validation.

and assembly variability. The data acquired with the first three pen prototypes were employed to calibrate the algorithm, while the rest of the data from seven pens was used to evaluate the stroke segmentation algorithm.

For the segmentation into strokes, given the pen force signals F, as a first step, bias from the force signal was removed; this is particularly important because the force bias may differ between diverse prototypes due to various factors (e.g. the tolerances of the 3D printed components, the ink refill replacement by the user, the curvature of the load cell wire during assembly). Since signal processing is fully automated during unsupervised pen usage, an automatic baseline removal was preferred. To do so, a value corresponding to the mode of the median filter computed over a 50-sample time window was first subtracted from the signal. After that, a baseline estimation and denoising algorithm was run using sparsity (BEADS)[2]. Finally, all the values

[2]from Laurent Duval (2020). BEADS Baseline Estimation And Denoising with Sparsity. https://www.mathworks.com/matlabcentral/fileexchange/49974- beads-baseline-

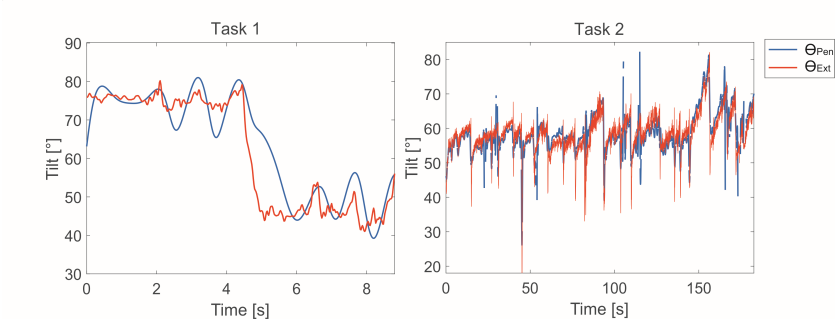

Figure 4.10: Experimental set-up for the tilt angle validation.

below an experimentally fixed threshold were cropped to zero; this zero-cropping threshold was chosen to optimise the agreement between the stroke identification obtained from the pen (first three pen prototypes) and the reference force signal.

As for the stroke segmentation of the reference force signal, the mean value of a manually selected non-writing tract was considered as baseline; after baseline removal, the strokes segments were defined as the non-zero parts in the signals.

A total number of 452 strokeswas detected in the pen signals. Through a Bland-Altman plot analysis [144], the agreement between the strokes duration obtained from the smart pen (last seven prototypes) and from the external force signal was evaluated. The linear regression between the strokes duration obtained in the two cases was also computed.

Results

A visual representation of the stroke segmentation in a 30-second long force signals is shown through horizontal tracts placed below the overlapped time series in Fig. 4.11. The Bland-Altman plot for the agreement between the stroke segmentation obtained from the pen writing force signals F and the reference signals from the external sensor F_{ext} is reported in Fig. 4.12; the mean stroke duration (in seconds) is presented in the x-axis, while the difference between the stroke duration between F and F_{ext} is in the y-axis. A total number of 452 strokes were detected, with a mean duration of 0.11 ± 1.96s. As can be observed from Fig. 4.12, there is agreement between the strokes duration obtained from the two signals: all data points except three were located inside the confidence interval boundaries of agreement, with no trends in the point distribution. The outlier points indicated the rare cases in which a large discrepancy between stroke duration was detected. These cases may correspond to the events in which the minimum values

estimation-and-denoising-with-sparsity

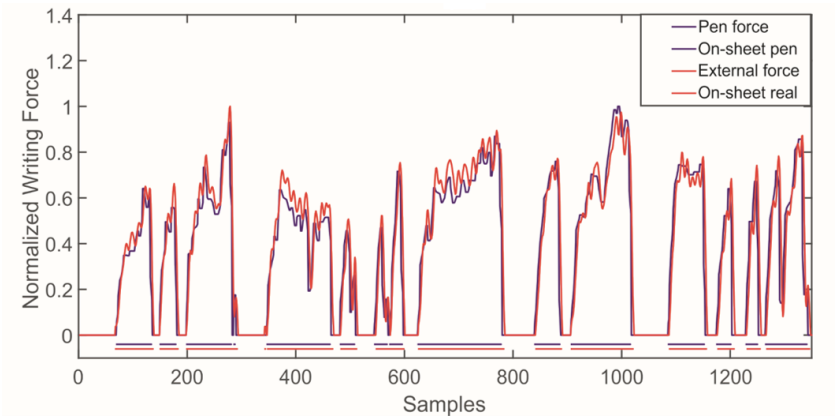

Figure 4.11: Aligned and overlapped writing force signals acquired with the pen (blue) and with the external load cell used as reference (red). Below the signals, the coloured segments (blue for the pen and red for the external load cell) indicate the detected strokes.

of pen measured force are just above the zero-cropping threshold, therefore the stroke detection algorithm does not divide the non-zero force tracts into different strokes. However, the zero-cropping threshold was set in order to minimise the number of outliers.

The linear regression between the stroke duration obtained from F and F_{ext} reported an R_2 score of 0.99.

4.3.5 Discussion

This part of the work presented the development, validation, and testing of a smart ink pen instrumented with force and motion sensors, designed for the quantitative and ecological assessment of daily-life handwriting.

The device was designed to autonomously acquire, store, and dispatch data related to the writing gesture. The pen is instrumented with an inertial measurement unit and a miniaturised load cell, so that the combined analysis of data collected through these sensors allows the extraction of relevant handwriting and tremor indicators. The smart object features both internal storage capacity and BLE connectivity, so that the collected data can be streamed to a remote device either in real-time, or offline after being stored onboard. To maximise usability and transparency, the pen functioning is self-operated through movement detection or BLE connection request, therefore no activation button for the end-user is needed; in addition, the sensor data download and processing procedures are automated.

The greatest accomplishment of the devised technology is that it enriches a traditional ink pen with the ability of achieving quantitative reliable handwriting assessment, thus combining the advantages of the digitizing tablet technology with the natural 'feel', the ease of use, and the ecological valid-

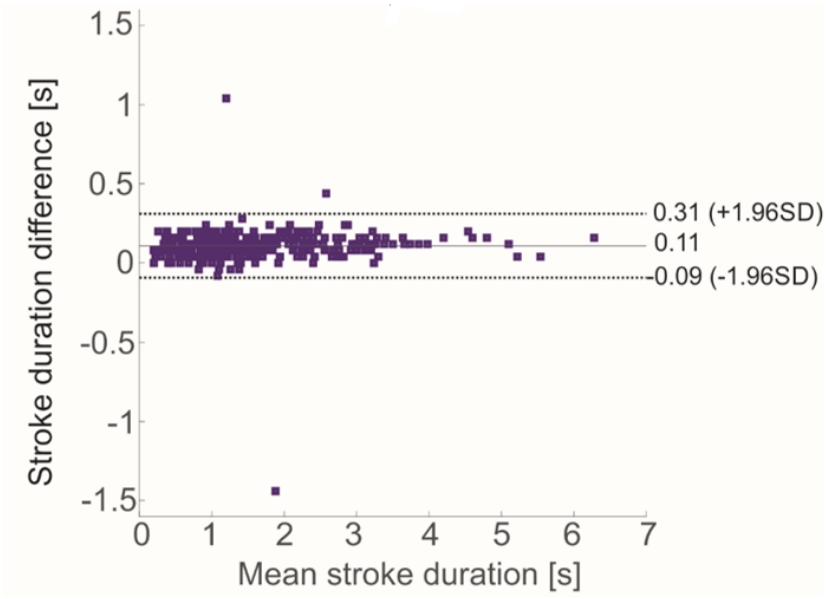

Figure 4.12: Bland-Altman plot with the duration of the strokes detected with the pen and the external load cell signals.

ity of the traditional pen-and-paper approach. These latter characteristics are crucial desires for the end-user, especially when dealing with the elderly population. Indeed, simplicity, intuitiveness, and transparency are key requirements to increase the acceptance and reduce the technological anxiety that often characterises the elderly population [145]. This is even more true when the device is envisaged within the framework of continuous home-based monitoring, as in the case of the proposed technology; the smart ink pen was indeed developed within the European Movecare Project [12], which targets independent older adults living at home with the final aim of detecting early signs of cognitive and physical decline.

The devised smart pen was first successfully validated against gold standard references. As for the pen load cell, the static tip force measurements calibration using testing weights reported a coefficient of determination extremely close to 1 (0.99), confirming the linearity of the sensor even when constrained in a very tiny space that imposes a pronounced bending of the cable. Comparison of the pen force signal with the reference measurement obtained from an external load cell reported a very strong correlation (ρ_d= 0.96) also in dynamic conditions. The validated pen force signal was leveraged for the segmentation of the writing signals into strokes, a critical step of the data analysis process since the computation of important handwriting indicators relied on it. The validation of the segmentation procedure conducted by comparing the strokes duration obtained from the automatic

stroke detection algorithm and the reference segmentation was successful. Indeed, the Bland-Altman Plot showed a very good agreement between the two sets of measurements, which were included, for the most part, within the standard deviation boundaries; in addition, an almost perfectly linear relation ($R_2 = 0.99$) emerged from the regression analysis between the two sets of strokes duration. As for the pen inertial measurement unit, it was leveraged to estimate the pen tilt angle; the pen tilt was validated through the comparison with a reference angle obtained by means of an optoelectronic system, reporting strong correlation and low error.

4.4 Handwriting Indicator Reliability

This section includes the protocol and the analysis aimed at studying age-related changes in handwriting and tremor parameters [3].

4.4.1 Participants and Protocol

After validation, the smart pen was used to collect handwriting data on a population of healthy young and older adults, with the aim of testing the reliability of the writing and tremor indicators, and to study possible age-related differences in handwriting and tremor features.

The study included voluntary subjects from different age ranges: young adults (age $<$ 50 years old) and older adults (age $\leq$ 60 years old). The exclusion criterion was any diagnosis of neurological, vascular, or musculoskeletal disorder affecting the upper limbs.

To maximise the ecological validity of the protocol, a task mimicking daily-life writing was proposed, without constraining the writing modality or content. Participants were seated at a table and asked to use the instrumented pen with their dominant hand to write 7 lines of free text on a paper sheet [127]. Participants executed the task twice, with a between-trial break of at least 4 hours. To obtain an ecologically valid test-retest reliability, no constraints on the writing content or modality were imposed to the user.

The Ethical Committee of the Politecnico di Milano approved the study protocol (n. 10/2018).

4.4.2 Indicator Calculation

Data analysis was carried out with Matlab® R2020b (Mathworks®, Natick, MA USA). Starting from the stored data package defined in Sec, 4.2, a set of writing and tremor indicators are extracted for each subject. For the writing indicators, as a first thing, the pen tilt was computed over the entire writing

[3] This study has been published in the EMBC IEEE 2020 conference [105]

period, following the method presented in Sec. 4.3.3. The pen tilt during the writing gesture has been included in studies investigating handwriting in a number of conditions [146, 147, 141]. For this reason, the mean (*Tilt mean*) of the pen tilt signal was retained, not including the segments longer than 2 seconds which were considered pauses. To evaluate how pen tilt varied during the writing gesture, the coefficient of variation *Tilt CV* was computed.

After that, the signals were divided into strokes (Sec. 4.3.4). The stroke segments were defined as *on-sheet* (or on-paper), and the complementary non-writing segments as *in-air*, not including the segments longer than 2 seconds which were considered pauses. After that, since previous work reported irregular writing rhythm with age [129], the following handwriting temporal indicators were computed:

- *In-air time*, t_{air} [s]: the average duration of the non-writing segments, averaged over all in-air tracts during the writing task execution.
- *On-sheet time*, t_{sheet} [s]: the average duration of the on-sheet segments, averaged over all strokes during the writing task execution.
- *In-air/On-sheet time ratio*, $t_{a/s}$: the ratio between t_{air} and t_{sheet}.

The force signal of each single stroke was low-pass filtered at 4 Hz (4^{th}-order Butterworth filter) [14] and the following force-related indicators were computed:

- *Mean writing force*, $\overline{F}$ [N]: the mean normal force applied to the pen tip in each stroke, averaged over all strokes.
- *Number of changes in force*, NCF: the number of local minima and maxima in pen force signal F in a stroke, averaged over all strokes [14].

As for the acceleration signals, a 3-axial gravity compensation was carried out for each single stroke. A compensation factor was computed, for each axis separately, as the average of the low-pass filtered single-axis acceleration signal (4^{th}-order Butterworth - cutoff: 3.5Hz), and then subtracted from the signal itself [148]. After gravity compensation, the signals were low-passed filtered at 4Hz (4^{th}-order Butterworth filter) and the 3D acceleration signal was computed for all the strokes. As a measure of smoothness the following measure was calculated:

- *Number of changes in acceleration*, NCA: the number of local minima and maxima in the 3D acceleration in a stroke, averaged over all strokes.

To study tremor, the acceleration signals was used without distinguishing between on-sheet and in-air tracts. The acceleration signals were cut into

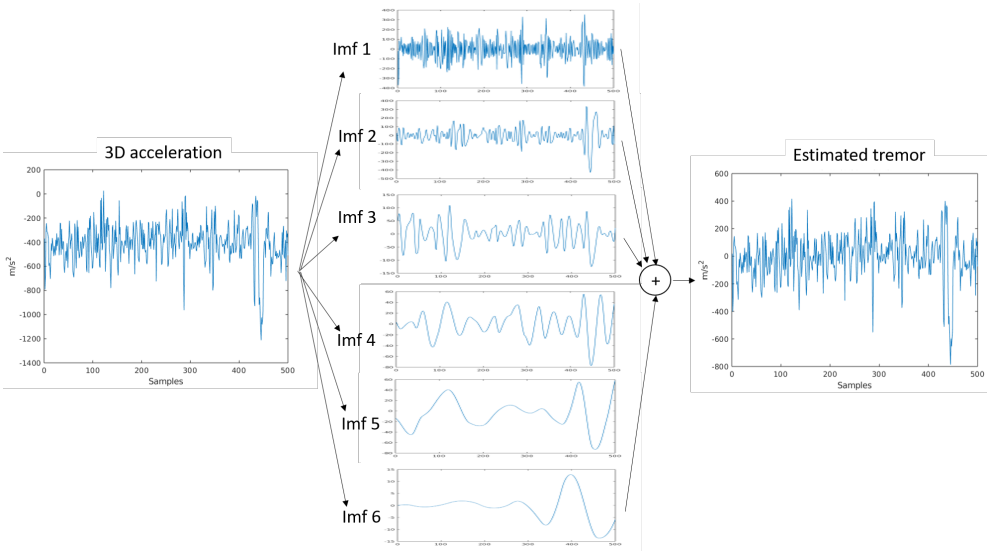

Figure 4.13: A visual example of the EMD of the 3D acceleration signal in the IMFs. The tremor signal is reconstructed by retaining the IMFs with a mean instantaneous frequency grater the 3Hz only.

500-sample segments [139]. The acceleration data over the 3 axes were firstly band-pass filtered at 0.5-24Hz (4^{th}-order Butterworth filter) [20]. The 3D acceleration signal was then computed as the module of the x-y-z components. For each segment, tremor was estimated using the Hilbert-Huang Transform (HHT), which consists in the Empirical Mode Decomposition (EMD), followed by the Hilbert Spectral Analysis [149]. The EMD-based filters decomposition is based on the local timescale characteristics of the data, since EMD does not have any general analytical formulation unlike other conventional fixed cutoff filtering techniques. For this reason, it was proven to be particularly suitable to study tremor in presence of voluntary motion [150], and it was thus preferred to the standard Fourier Transform to study tremor frequency components during our handwriting protocol. Each 500-sample segment of the 3D acceleration signal was decomposed in the intrinsic mode functions (IMFs) using EMD, with the Cauchy convergence criterion. Then, the IMFs whose mean instantaneous frequency was higher than 3Hz were retained only (because they could be associated to the tremor features [150], rather than voluntary motion). The remaining IMFs were summed to reconstruct the 3D tremor signal of each segment. A visual example of the EMD and the tremor signal reconstruction from the 3D acceleration is shown in Fig. 4.13. The tremor indicator were computed for each segment of the reconstructed tremor signal and then averaged. In particular, the following tremor indicators were computed:

- *Modal frequency*, f_{mod} [Hz]: the tremor characteristic frequency value

which corresponds to the highest peak in the Hilbert power spectrum of the tremor signal obtained from the IMFs [15], and has been shown to change in case of pathology [151].

- *Approximate entropy*, ApEn: the estimate of the entropy in the reconstructed tremor signal, a regularity statistic measure that quantifies the unpredictability of the fluctuations in a time series signal, and returns a value between 0 (high degree of short- and long-term predictability) and approximately 2 (completely random signal such as pure white Gaussian noise). For the computation of ApEn [152], m (the window length) was set to 2, and r (the similarity criterion) was set to 0.2xSD of the signal, as in [20]. Regularity of tremor has been shown to change with age and pathology [153].

After that, the Recurrence Quantification Analysis (RQA) was applied to the tremor data; it is a nonlinear data analysis method that quantifies the number and duration of recurrences of a dynamical system presented by its phase-space trajectory [154]. Two-dimensional binary maps - recurrence plots (RPs) - are computed for each tremor signal to visualise the recurrent behaviour of the phase-space trajectory of dynamical systems. The following settings were adopted [154]: i) the delay was estimated with the Mutual Information method algorithm; ii) the embedding dimension was estimated with the false nearest neighbour (FNN) chaotic algorithm [155]; iii) the critical radius was set to 20% of the maximum distance (Euclidean distance matrix). From these maps, the following indicators were obtained to describe the complexity in handwriting [139]:

- *Recurrence ratio*, RR [%]: it expresses the self-similarity of the tremor time series during handwriting.
- *Determinism*, DET [%]: an index of the degree of determinism, that expresses the predictability of the signal.

The higher these indicators, the lower the tremor complexity [139].

A schematic representation of the workflow is represented in Fig. 4.14 for the indicators calculation.

4.4.3 Statistical Analysis

Methods

Statistics was run using RStudio version 1.2.5033 (RStudio Inc., Boston, MA). Significance level was set at 5% for all tests.

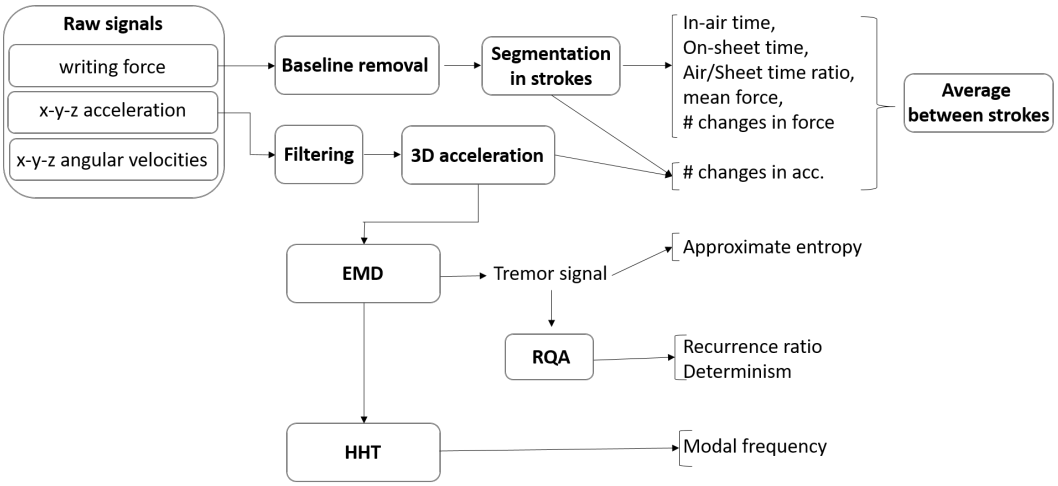

Figure 4.14: A schematic representation of the indicators calculation. The writing force signal is pre-processed using baseline removal and stroke segmentation. The averaged in-air and on-sheet duration among all the strokes are considered as the temporal indicator: in-air time, on-sheet time and the ratio between them. The mean force value and the number of changes in force are computed within each stroke and averaged among all the strokes. The number of changes in acceleration is calculated from the 3D acceleration signal within all the strokes an then averaged among all the strokes. Empirical mode decomposition (EMD) is used to estimate the tremor signal, from which the approximate entropy is obtained and the non linear quantities (recurrence ratio and determinism) are calculated applying recurrent quantification analysis (RQA). At last, the modal frequency is calculated as the highest peak in the power spectrum obtained using the Hilbert-Huang transform (HHT).

The goals of our statistical analysis were: i) to check the reliability of the computed indicators in a test-retest design, and ii) to study possible significant variations of these quantities with age in healthy subjects [4].

Data acquired in the two sessions were used to evaluate the relative and absolute reliability of the computed indicators, for young and older adults, separately. To do so, for each continuous indicator, a Lilliefors Test was first conducted to verify normality, and a paired t-test to assess the absence of systematic bias between the measurements in the first and in the second session of the test. The reliability was evaluated by computing the intra-class correlation coefficients (ICC 2-way mixed-effects model, absolute agreement), an index ranging from 0 to 1; ICC values of 0.5, 0.75, and 0.9 indicate moderate, good and excellent reliability, respectively [157]. Absolute reliability was assessed computing the standard error of measurement (SEM), as in Eq. 4.4:

$$SEM = SD \times \sqrt{1 - ICC}, \tag{4.4}$$

[4]This study has been published in the GNB 2020 conference [156]

where SD is the standard deviation of the indicator in the test-retest. Given the SEM, the measurement error was estimated through the Minimal Detectable Change (MDC) calculated as in Eq. 4.5:

$$MDC = SEM \times 1.96 \times \sqrt{2}, \tag{4.5}$$

where 1.96 is the z-score associated with the 95% level of confidence, and the square root of 2 reflects the additional uncertainty introduced by using scores based on measurements made at two time points [158]. On the other hand, for the two count indicators (NCA and NCF), test-retest concordance between the two test repetitions was investigated by computing the Kendall's Tau [159].

As for the second aim, a non-parametric statistics was chosen after verifying, with a Lilliefors test, that not all indicators were normally distributed. For each indicator, possible between-group differences were studied using the Kruskal-Wallis Test, followed by Wilcoxon pairwise comparisons with Bonferroni correction in case of significance.

Results

A total of 43 subjects was recruited. The young adults group (YY) included 20 subjects [Age 28.5 (mean) ± 3.6 (standard deviation) years old; Sex: 10 Male and 10 Female; Dominant hand: 19 Right-handed]; the older adults group included 23 subjects [Age 73.4 ± 8.9 years old; Sex: 8 Male and 15 Female; Dominant hand: 22 Right-handed]. To study how age affects handwriting, the subjects in the older adults group were divided into 2 additional based on their age: the subgroup of middle-old adults (EY) included 12 subjects with an age between 60 and 69 years old [Age 66.4 ± 2.1 years old; Sex: 6 Male and 6 Female; Dominant hand: 12 Right-handed]; the second subgroup (EE) was composed by 11 old adults with an age above 70 years old [Age 81 ± 6.9 years old; Sex: 2 Male and 9 Female; Dominant hand: 10 Right-handed]. The choice of separating subgroups in the population over 60 enables a more accurate portrayal of significant life changes. The 70 years old threshold was based on published studies showing that the relationship between age and handwriting movements is likely to be non-linear with the greatest decline in age-related motor function occurring after the age of 70 [147, 141].

For the reliability, Tab. 4.1 presents the reliability results for both young and old adults. The test-retest reliability results on the older adult group were computed from the data of 11 subjects who performed the protocol twice. As shown in Tab. 4.1, excellent or good reliability emerged for all the indicators, for both young and older adults, except for f_{mod} for the old adults, which showed moderate reliability (ICC = 0.68).

RELIABILITY OF THE HANDWRITING AND TREMOR INDICATORS

	YOUNG ADULTS			OLD ADULTS		
Indicator	**Test-retest reliability**			**Test-retest reliability**		
CONTINUOUS	**p-value**	**ICC (95% CI)**	**MDC (SEM)**	**p-value**	**ICC (95% CI)**	**MDC (SEM)**
Tilt_mean [°]	< 0.001	0.96 (0.90-0.98)	2.24 (0.81)	< 0.001	0.99 (0.98-0.99)	2.37 (0.85)
t_{air} [s]	< 0.001	0.93 (0.83-0.97)	0.04 (0.01)	0.001	0.88 (0.57-0.97)	0.08 (0.03)
t_{sheet} [s]	< 0.001	0.96 (0.9-0.98)	0.11 (0.04)	< 0.001	0.96 (0.87-0.99)	0.11 (0.04)
$t_{a/s}$ [a.u.]	< 0.001	0.91 (0.78-0.96)	0.12 (0.04)	< 0.001	0.95 (0.81-0.98)	0.17 (0.06)
$\bar{F}$ [N]	< 0.001	0.88 (0.71-0.95)	1.5 (0.54)	< 0.001	0.96 (0.88-0.96)	2.02 (0.73)
f_{mod} [Hz]	< 0.001	0.80 (0.51-0.92)	0.53 (0.19)	< 0.001	0.68 (0.11-0.91)	1 (0.36)
ApEn [a.u.]	< 0.001	0.90 (0.74-0.95)	0.03 (0.01)	0.002	0.85 (0.47-0.96)	0.08 (0.03)
%RR [a.u.]	< 0.001	0.82 (0.55-0.93)	0.1 (0.04)	0.01	0.77 (0.21-0.94)	0.17 (0.06)
%DET [a.u.]	< 0.001	0.88 (0.72-0.95)	0.06 (0.02)	< 0.001	0.89 (0.60-0.97)	0.06 (0.02)
COUNT	**p-value**	**Kendall's Tau**		**p-value**	**Kendall's Tau**	
NCF [#]	< 0.001	0.85	-	< 0.001	0.88	-
NCA [#]	< 0.001	0.83	-	0.002	0.81	-

Table 4.1: Results of the statistical analysis for the test-retest on 20 young adults and 11 older adults. CI=Confidence Interval. For continuous indicators, test-retest reliability was investigated with the ICC, while Kendall's Tau was computed for count indicators.

As for the age-related changes in handwriting, Tab. 4.2 reports the results of the non-parametric statistical analysis carried out to test the age effect on the handwriting indicators. Significant between-group differences were found for in-air time (t_{air}), in-air/on-sheet ratio ($t_{a/s}$), number of changes in force (NCF), approximate entropy (ApEN), recurrence ratio (RR) and determinism (DET). Pairwise comparisons highlighted significant changes for: the couples YY-EY and YY-EE for t air (increase with age); the couples YY-EE and EY-EE for $t_{a/s}$ (increase with age), RR and DET (increase with age); the couple YY-EE for the ApEn (decrease with age) and *tiltCV* (decrease with age) indicators; the couple EY-EE only for NCF (decrease with age); and the couple YY-EY only for *tiltmean* (decrease with age).

4.4.4 Discussion

After a successful validation of the pen sensors and algorithms, the smart object was tested on young and older adults with the twofold aim of testing the reliability of the handwriting and tremor indicators, and their sensitiv-

AGE-RELATED CHANGES IN HANDWRITING

Indicator	YY (Median	IQR)	EY (Median	IQR)	EE (Median	IQR)	x^2	p-value	YY vs EY (p-value)	YY vs EE (p-value)	EY vs EE (p-value)
Tilt_mean [°]	66.5	9.8	59.1	4.7	60.3	9	10.1	**0.006**	**0.005**	0.59	0.18
Tilt_CV [°]	0.13	0.04	0.16	0.06	0.11	0.03	9.1	**0.01**	0.30	**0.028**	0.07
t_{air} [s]	0.28	0.1	0.40	0.2	0.75	0.46	18.2	**1.1e-4**	**0.013**	**1.4e-4**	0.11
t_{sheet} [s]	0.55	0.25	0.59	0.14	0.63	0.33	0.88	0.64	-	-	-
$t_{a/s}$ [a.u.]	0.47	0.25	0.54	0.17	0.64	0.18	11.8	**0.003**	0.76	**0.002**	**0.04**
$\bar{F}$ [N]	1.87	1.62	1.77	1.50	1.82	1.52	2.8	0.24	-	-	-
NCF [#]	4	3	4	1.25	3	1	6.5	**0.04**	1	0.25	**0.02**
NCA [#]	3	2.25	4	1.5	5	2.5	0.89	0.64	-	-	-
f_{mod} [Hz]	5.2	0.44	5	0.52	4.75	1.22	3.2	0.2	-	-	-
ApEn [a.u.]	1.14	0.09	1.06	0.11	0.9	0.19	16.1	**3.2e-4**	0.08	**1.8e-4**	0.095
%RR [a.u.]	0.39	0.11	0.3	0.1	0.49	0.08	12.9	**0.002**	0.14	**0.02**	**0.004**
%DET [a.u.]	0.74	0.13	0.72	0.15	0.87	0.04	13.5	**0.001**	1	**2.9e-4**	**0.017**

Table 4.2: Results of the non-parametric statistical analysis between YY, EY and EE groups. The first p-value reported (column 9) represents the p-value of the Kruskal-Wallis Test, together with the x 2 (column 8); for those indicators that reported a significant p-value in the Kruskal-Wallis Test, pairwise comparisons were conducted through the Wilcoxon Test (Bonferroni adjustment), and the related p-values are reported in columns 10, 11 and 12.

ity to distinguish between age groups [5]. Concerning test-retest reliability, results were excellent, with values largely above the 0.75 threshold of good relative reliability [157], with the exception of the modal frequency for the old adult group, which showed moderate reliability. These high values are even more striking if one considers that the writing content differed between the two trials, importantly suggesting that the reliability of the indicators is independent from the writing content. The minimal detectable change was also computed to estimate whether a change between the user's repeated tests represents random variation or a true change in performance. This measure is extremely important to discriminate real changes in the values of the indicators when monitoring users over time, and it is thus crucial for the user's longitudinal monitoring to highlight relevant deviations from the standard performance. The reported measurement errors were very low.

Concerning the ability of the writing and tremor indicators to distinguish between different age groups (young adults YY, middle-old adults between 60 and 69 years old EY, old adults over 70 years EE), our results are mostly in line with previous literature.

As for the temporal handwriting measures, significant age-related changes emerged for the in-air time and the in-air/on- sheet ratio indicators, which

[5]The research has been published in the IEEE Transactions on Instrumentation and Measurement Journal [160]

increased with age. Our findings confirm previous literature investigating writing copying tasks using digitizers, which reported an increasing trend for these indicators from healthy young to older adults [129], and from healthy older adults to older adults with cognitive decline [10]; moreover, the assessment of in-air time during handwriting was reported to have a major impact on disease classification accuracy in PD [128]. On the other hand, our data showed how the On-sheet time indicator was quite constant over different age groups. As for the indicators related to the handwriting force (Mean writing force), no age differences emerged. For this indicator, contrasting results emerged in previous work analysing writing copying tasks using digitizers: while no significant association between writing pressure and the frailty phenotype was found by Camicioli et al., [130], some studies reported that the writing pressure decreases with age [161], and age-related cognitive pathology [10]. The lack of age-related differences in writing force emerged from our data is unlikely to depend upon the force resolution of the pen load cell. Indeed, from the static calibration, a force measurement resolution of 0.02N was obtained, which is two orders of magnitude smaller than the MDC value found for the force signal (equal to 1.5N). On the other hand, our data showed an effect of age for the number of changes in force within a stroke, which was significantly decreased for the very old group (EE). Our finding confirms previous literature examining pen-and-paper writing, which reported a more uniform pen pressure for older writers [8].

Concerning tremor, no differences due to age emerged for the Modal Frequency. While this indicator was shown to be affected by the presence of neurological conditions (e.g.: PD) [151], a clear effect of age was not consistently shown in previous literature, which mainly investigated resting or postural tasks [139, 20, 15, 153]. On the other hand, our data revealed a neat age effect on nonlinear tremor acceleration characteristics. Our results showed a decrease with age of approximate entropy during handwriting, meaning that more repetitive and predictable tremor oscillation components characterised the older age groups. As for this indicator, previous literature showed a clear decrease in PD patients [153], and a slightly less evident decreasing trend for older individuals, compared to younger adults, in postural rather than in resting tasks [20, 15, 153]. In line with the approximate entropy results, the indicators related to the recurrence quantification analysis, recurrence ratio and determinism, presented a significant increase for the very old group (EE), confirming once again the augmented predictability of the tremor characteristics in older writers. Also for these indicators, previous work revealed a marked increasing trend following neurological conditions (PD) [139], while no important changes simply due to older age were consistently demonstrated, at least during postural tasks [139].

To sum up, the results of the writing indicators obtained during free text writing with our pen, especially those related to temporal measures, are in

line with previous literature investigating writing copying tasks with digitizing tablet technology. This result supports our choice of not constraining the writing content or execution in order to increase the ecological validity of the protocol. As for tremor, the nonlinear acceleration characteristics examined in our study while writing with the smart pen present a more marked effect of age, compared to previous work investigating mostly postural tasks [139]. This finding suggests that the study of tremor during more complex activities entailing a blend of cognitive and fine motor skills, such as handwriting, is more effective in bringing out important differences due to ageing, compared to more simple postural or resting tasks typically investigated in previous work. To this end, the proposed technology, with its combination of force and motion sensors, is key since it allows the simultaneous study of writing and tremor characteristics during handwriting tasks. This important advantage of the smart pen paves the way toward fruitful applications of the current technology in the field of Parkinson's disease. Indeed, signs of the disease include not only tremor, but also a series of handwriting abnormalities grouped under the term of 'PD dysgraphia' [162], which supports the study of handwriting as a pre-symptomatic neurobehavioural biomarker of PD [163]. In this framework, it is clear that a technology that allows the combined study of handwriting and tremor features perfectly suits the current needs of the neurological research field. Indeed, the importance of studying handwriting is not restricted to Parkinson's Disease, but can be extended to a variety of other neurological disorders, including dyskinesia [164], Huntington's Disease [126], and Multiple Sclerosis [165], not only to support the diagnosis process, but also to quantify the severity of clinical signs over time and to monitor and manage the risks associated with medications [164]. In addition, a technology that allows quantitative, simple, and ecologically valid evaluation of handwriting finds potential and interesting applications also in the youngest population, since handwriting and text production skills assume a central role in the children's development process. In this framework, handwriting difficulties are common in a number of childhood disorders, including, dysgraphia [166, 167], dystonia [168], and attention deficient hyperactive disorder [169].

4.5 Usability and User Experience

Methods

The smart pen was included in the eHealth platform for home monitoring and assistance of the European project MoveCare (H2020, GA no. 732158) [12], targeting independent older adults. Against this background, the ink pen was considered as an IADL monitoring device. The pen was designed to autonomously collect and upload handwriting data when used, with the minimum action required by the user beyond simply write with the pen

and put it in charge. The project was finalised to an ecological momentary assessment in which the platform was tested in the real-world scenario with 25 community-dwelling older adults (65 or more years old) living alone. In particular, participants will have to satisfy the following inclusion criteria: i) age $\geq$ 65; ii) living alone (with assistance in activities of daily living $\geq$ 1 hour/day); iii) Mini-Mental State Examination (MMSE) $\geq$ 26; iv) pre-frail (Fried Scale = 1-2) people or robust (Fried Scale = 0) individuals who experience loneliness (Revised UCLA Loneliness Scale $>$ 35) and depression (Geriatric Depression Scale (GDS) $>$ 9).

The pilot study was organised in two rounds (round 1 and 2), each of them lasting 3 months. Some of the participants (14 'RMHS' subjects) were recruited in Spain (Badajoz), the others in Italy (Milan). People of the Italian group people were recruited among the local association of older volunteers A.N.T.E.A.S (Associazione Nazionale Terza Eta' Attiva per la Solidarieta') and the patients afferent to the Day Hospital and the Ambulatory of the Geriatric Unit of Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico (PCL). In addition, 4 residents of the Heliopolis Centre of the Korian group were included. In the final pilot test, usability and acceptance of the various components of the eHealth platform were investigated in two main ways: i) through users' feedback collected using questionnaires; ii) by measuring the amount of time the devices/modules were used.

Results

The usability measurement of the smart pen for each subject in the first and second round are showed in Fig. 4.15. The upper panel shows a bar-plot that counts the number of data entries for each subject, e.g. the number of time the subject has used the pen during the pilot study. The lower panel, instead, shows the bar-plot counting the amount of time (in minute) each subject has used the pen, for each round separately. Different colours indicate pilot users recruited in Italy by PCL (in green) in Italy by Korian (in red) and in Spain by RMHS (in blue). For all users, dark colours indicate users recruited in round 1, and light colours users recruited in round 2. Among the monitoring systems included in the MoveCare platform, the smart pen resulted used by the most of the recruited subjects.

Questionnaires about the smart pen included 4 5-point Likert scale questions (using 1 for 'totally disagree' and 5 for 'totally agree'). The questions and the relative mean and median score obtained are reported in Tab. 4.3 The scores obtained by the Smart objects (ball and pen) were very high, especially the Smart pen. 90% of the participants rated question one *using the smart pen was easy* positively and the 86.4% rated question two *I found it easy to charge the smart pen' positively.* The bar chart with the frequency

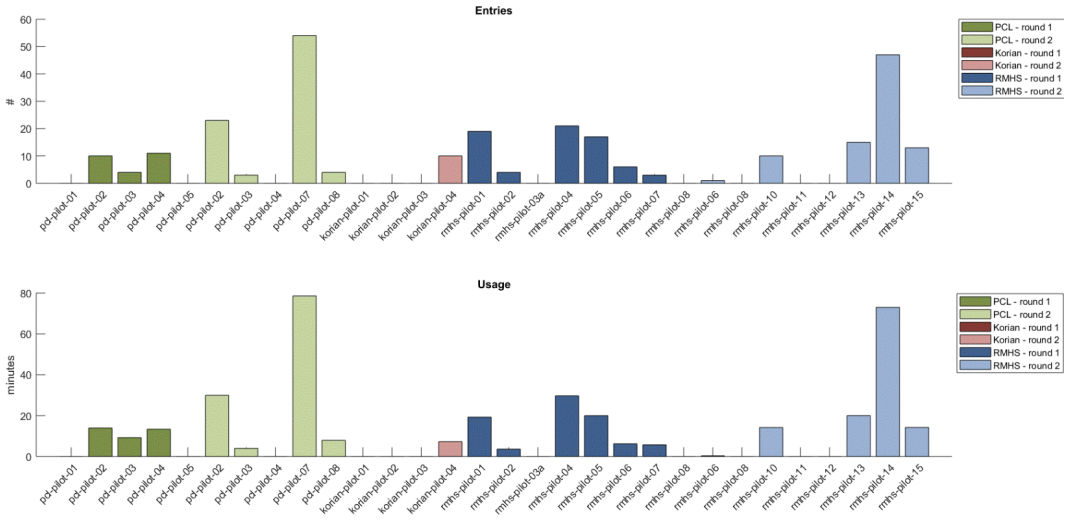

Figure 4.15: The number of valid entries (upper panel) and the minutes of usage (lower panel) of the whole users involved in the final pilot testing.

Question	**Mean**	**Median**
Using the pen was easy	4.68	5
I found it easy to charge the pen	4.68	5
I think monitoring handwriting through the use of an instrumented ink pen is useful	4.50	5
If I have the possibility, I would frequently use the smart pen in the future	4.23	5

Table 4.3: Usability questions with the relative mean and median score value.

distribution of the given answer for questions one and two are showed in Fig. 4.16 . Statistically significant differences have been identified for question one 'Using the smart pen was easy' in the comparison by country ($p = 0.009$): the proportion of Spanish participants who valued this question as fully agree is significantly higher than that of Italian participants (100.0% vs. 44.4%).

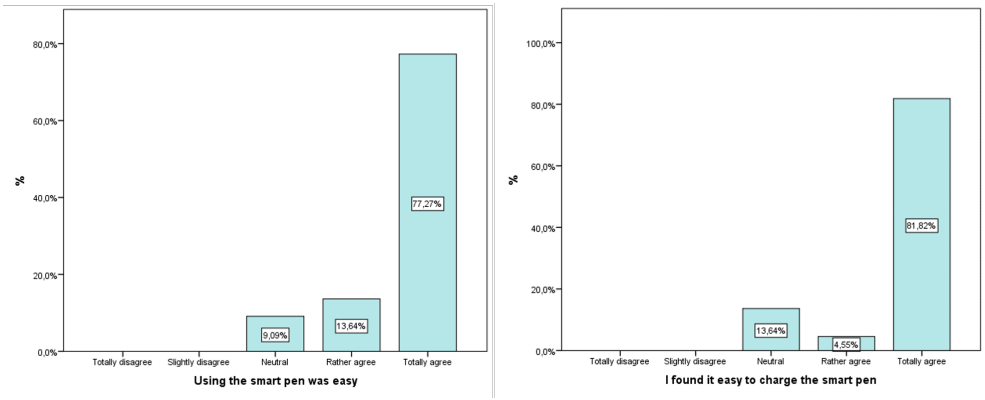

Figure 4.16: Frequency distribution of given answer for the questions one (left panel) and two (right panel).

Chapter 5

Anomaly Detection in Uncontrolled Handwriting

As described in the previous chapter, experiments revealed that the smart ink pen was a suitable and reliable tool for the ecological collection of quantitative information about the handwriting activity. In the home-based monitoring scenario, any evaluation of the health status through the assessment of the handwriting performance cannot rely on standard protocols. Indeed, the analysis of the handwriting indicators, in this research, has been conducted using data collected by asking subjects to perform common daily-like writing tasks (e.g. short free text), without any bind on the writing content or modality [1].

Although reliable measurements could be obtained using the pen, the subject's handwriting could not be evaluated by simply comparing the indicator with those values which, in previous studies, have been associated to certain conditions of decline. The ecological data acquisition modality, required in the home monitoring context, increases the noise and the data variability. So measurement may have erratic fluctuations and unforeseen behaviours. This inevitably implies that the signs of a possible decline are not easily attributable to already known values or trends of the measures. Consequently, the sign to look for is the presence of anomalies in the data.

Anomalies can be defined as outliers or significant trend changes in the data and, in the case of multidimensional data, they can be also represented by particular patterns of indicators. Anomaly detection (AD) techniques basically consists in finding those outliers, trends, or patterns in the data that are usually related to some problems or malfunctioning events [170]. In free handwriting, AD could be employed to identify some variations in the writing data which might suggest an alteration of the physical or mental status of the subject. In the case of older adults and frail individuals, it

[1] See IEEE-EMBC paper [105]

might be related to a pathological decline in capacity.

Given the uncontrolled condition in which handwriting data will be acquired in the home monitoring scenarios, the quantities to observe should be accurately chosen. In addition, some consideration needs to be done.

- The monitored subject should be in a known initial health status and such status is expected to undergo variations in the middle-long period. For example, a target subject can be a healthy older adult, since the old age is known to be a risk factor for conditions such as frailty, or a pathological individual, in the early stage of the disease, who has the risk of getting worse.
- A time interval in which the health status of the subject is not expected to undergo significant variations should be identified. More precisely, changes in handwriting should not be visible within that period. For an healthy subjects, this period has been observed to be about 5 years [8]. It is expected to decrease in non-healthy individuals. Within this interval, the characteristic of the subject's handwriting can be assumed as time-invariant. Therefore, time-invariant measurements can be found and monitored.

At this point, anomalies can be interpreted as significant changes of those variables that were expected to be stationary [2].

The quantitative analysis of handwriting offers a set of indicators which can be treated as the time-invariant variables to be monitored. The monitoring of a subject's handwriting parameters can follow two different approaches: i) the subject-, or ii) the population-specific methods. In the subject-specific monitoring, anomalies are searched in the time series generated by sampling the subject's handwriting indicators over time. This method has the advantage to be more precise in the detection of decline for the subjects, being the inter-subject variability very high. On the other hand, it requires a long data collection on the same subject in order to test and evaluate the methodology. The population-based approach instead, finds the common patterns for the handwriting indicators in homogeneous population groups. For example, in healthy groups of individuals within the same age range. Here, anomalies can be searched in the subject's handwriting indicators which result associated a group different from the one of the subject (for example an older or a pathological group). This second approach might not be as accurate as the first one, but it can be implemented and tested more easily, since the variables to be monitored are not time-dependent.

The rest of this chapter will show the implementation of a population-based anomaly detection technique using the handwriting indicators ob-

[2] In the *weak* sense.

tained with the use of the smart pen. It is organised as follows: Section 5.1 presents the experimental protocols, the data processing and the classification algorithms used in the age-groups classification with the relative results. Section 5.2 shows the methods and the results of the application of the classification methods in the discrimination between healthy and Parkinson's patients. At last, Section 5.3 discusses the results.

5.1 Age-Group Classification

Here, the ability of the handwriting indicators in the classification of four age groups of healthy subjects was investigated performing two types of unconstrained writing tasks.

5.1.1 Participants and protocol

A total number of 80 healthy participants ranging from 20 to 90 years of age were involved. Any diagnosis of neurological, vascular, or musculoskeletal disorder on upper limbs was the exclusion criterion. Subjects over 65 years were admitted after verifying a Mini-Mental State Examination (MMSE) [171] score higher than 25. All participants were divided into four groups defined by age: group YY between 20 and 39 (12 males, 8 females, mean age 27.4±2.4), group EY between 40 and 59 (12 males, 8 females, mean age 57.7±6.28), group EF between 60 and 69 (10 males, 10 females, mean age 65.45±2.2), and subjects older than 70 (6 males, 14 females, mean age 80.2±7) were included in group EE. Each group included 20 subjects.

All subjects wrote a free text (*Text*, up to 10 lines) and a grocery list (*List*, up to 8 words). The tasks had no specific constraint to to make the task very similar to a daily life-like writing activity. The Ethical Committee of the Politecnico di Milano approved the study protocol (n. 10/2018).

5.1.2 Calculation of the handwriting indicators

A set of 14 parameters, related to the handwriting kinematics and dynamics, and to tremor, were extracted from raw data collected during each of the two writing tasks. The calculation was implemented in Matlab® R2020b (Mathworks®, Natick, MA USA) [3]. The following handwriting indicators were computed:

- *In-air time, On-sheet time* and *Air-Sheet time ratio.* The temporal parameters were obtained by analysing the writing force signal. The time during writing was divided into strokes, defined as the "writing

[3] See Lunardini et al. [160] for a detailed description

segments" in which the pen tip was in contact with the paper surface (non-zero force tracts). Then the averaged stroke length within a writing task was considered as mean OnSheet time (T_{sheet}). Unlike for the non-writing segments, the averaged zero-force tracts length was treated as the mean InAir time (T_{air}). InAir time intervals longer than 2 seconds were excluded as they were treated as pauses. The ratio between the latter and the former quantity was defined as the AirSheetR ($T_{a/s}$). The temporal parameters has been shown to describe a more irregular rhythm with age [129].

- *Mean value and variation coefficient of the tilt angle.* The pen inclination angle was computed within the writing task, excluding pauses, applying the sensor fusion algorithm (reported in [160]) to the inertial measurements. Previous studies included the pen inclination along the writing gesture to characterise handwriting in different conditions [146, 141]. The mean value ($Tilt_{mean}$) and the coefficient of variation ($Tilt_{cv}$) of the tilt angle signal were considered during writing. An angle equal to 90° indicated the straight vertical positioning of the pen and 0° represented the complete inclination, with the pen placed horizontally.

- *Mean value and number of changes in force.* The mean writing force (F) was calculated by averaging the force signal over all the strokes detected in the writing task. The information about the force variation was obtained by computing the mean number of changes in force (NCF) by counting the local maxima and minima within a stroke, averaged over all the strokes. Force and force variations have been noticed to change with age in handwriting [14].

- *Number of changes in acceleration.* The number of local minima and maxima in the 3D acceleration in a stroke was averaged over all strokes and was considered as the mean number of changes in acceleration (NCA). This quantity was associated to the smoothness of the writing movement and was observed to decrease with age [8].

- *Modal frequency and root mean square (RMS).* To extract information about tremor, the linear acceleration recorder was cut over a writing task signals in segments of 500 samples each [139]. The power spectrum was computed for each segment using the Hilbert-Huang transform (HHT) [149], which has been preferred in literature for the study of voluntary tremor to the standard Fourier transform [150]. Then, the mean modal frequency (f_{modal}) was obtained by averaging the frequencies of the highest peak in the power spectrum over all the segments [15], and the mean RMS by averaging the root mean square of the power spectrum over all the segments.

- *Approximate entropy.* The linear acceleration signals of the writing task, band-pass filtered at 0.5-24Hz (4^{th} order Butterworth filter), were used to estimate the approximate entropy ($ApEn$) as in [20]. The entropy value (ranging from 0 and 2) measured the unpredictability of the acceleration signals, which can be influenced by the higher or lower regularity of the tremor components. Entropy has been measured to decrease with age and pathology [153].

- *Recurrence ratio and percentage of determinism.* The nonlinear characteristics of tremor were quantified by performing the recurrence quantification analysis (RQA) to the acceleration signals. As in [139], the recurrence ratio (RR) was retained to measure the tendency of the tremor dynamics to express repeated pattern in time and the percentage of determinism (DET) to estimate the predictability of the the gestures during handwriting.

5.1.3 Dataset description

According to the protocol described in section 5.1.1, D_T was defined as the free text dataset composed of 80 samples and 15 attributes, i.e. the 14 indicators and the group label. In the same way, D_L was defined as the grocery list dataset. Finally, D_{TL} was the dataset composed of 80 samples 29 attributes, i.e. the indicators of both the tasks plus the group label. Given the set of the four ordered age intervals $A = \{YY, EY, EF, EE\}$ and the set of writing tasks $W = \{T, L, TL\}$, D_w^a was defined with $a \in A$, $w \in W$ as the dataset composed by 20 samples and computed from the group a over task w.

5.1.4 Classification tasks

The ability of handwriting to discriminate subjects belonging to different age groups was investigated. The handwriting indicators measured high-level phenomenons that emerged from the underlying complex processes of ageing. Therefore, machine learning (ML) classification techniques were used to account for the multivariate and non-linear nature of the problem. Two different ML algorithms were chosen to compare different classification logic. Logistic regression was used to set a baseline performance measure, since it is one of the simplest and most used linear classifiers. The second one was a more recent boosting algorithm named Catboost [172]. This algorithm is known to achieve remarkable performance while avoiding data overfitting, even with small datasets.

Since one of the most fruitful monitoring applications of our analysis is the detection of age-related abnormalities in the handwriting data, the attention was more focused over binary classification tasks to discriminate between age groups. In details, two pools of classification tasks were computed: the

first one is between adjacent groups (by age ranges), i.e. YYvsEY, EYvsEF and EFvsEE. The second one is between YY, EY and the EE group. In all the tasks, the younger class was indicated with '0' and the older one with '1'. The first pool of tasks was aimed at evaluating the performance of the classifiers in discriminating between groups within close age ranges. Models which achieve high performance in those tasks are expected to be more sensitive to minimal changes in handwriting due to the age-decline process. For monitoring purpose, Precision is the most significant metrics since it measures how much the classifier is robust in the determination of the true positives. The second pool of tasks was rather aimed at assessing the stronger ability of the models to detect more relevant changes in more distant age-groups. As resumed in Fig. 5.1, for each experiment, data normalisation in the range [0,1] was applied. Then, samples were labelled with

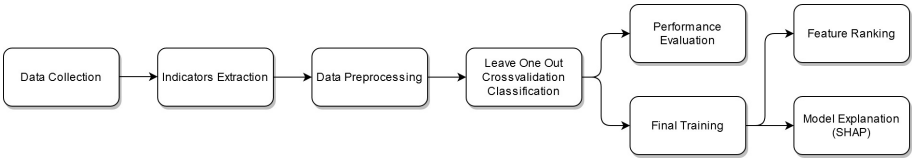

Figure 5.1: The data processing, classification and model explanation workflow.

1 for the older group, and 0 for the younger. In this way, the ML algorithms learned to predict the sample probability to belong to the right class. For each experiment the models were evaluated according to a wide set of classification metrics, i.e. Accuracy, Precision, Recall, F1 and Area Under the ROC Curve (ROC-AUC). To have less unbiased estimation of performance, both Logistic Regression and Catboost were evaluated with default parameters by the Leave-One-Out (LOO) Crossvalidation with early stopping set to 20 epochs.

5.1.5 Model explanation techniques

A model explanation technique was used to overcome the limitations of the black-box nature of the classification algorithms and to gain precise information about the models' decisions, i.e. the importance and the role of the handwriting indicators in the prediction of the subject's age group. The SHAP tool was used [173, 174], which is a model explanation library based on game theory which computes the Shapley values [175] of the features according to their impact on its predictions. For each trained model, it computes the Shapley Values [175] of the features according to their impact on its predictions. In a binary classification tasks, SHAP first computes the baseline prediction value, i.e. the mean value predicted by the model given the observed samples, then assigns a real number to weight each feature ac-

Table 5.1: Catboost scores evaluated through LOO Crossvalidation for age-groups Binary Classification.

	Accuracy			Precision			Recall			F1			ROC-AUC		
	Text	List	Text+List	Text	List	Text+List	Text	List	Text+List	Text	List	Text+List	Text	List	Text+List
YY vs EY	82.5	**85.0**	80.0	**84.2**	81.8	83.3	80.0	**90.0**	75.0	82.1	**85.7**	78.9	93.0	**96.8**	95.5
EY vs EF	**90.0**	82.5	85.0	83.3	**84.2**	81.8	**100.**	80.0	90.0	**90.9**	82.1	85.7	**98.0**	92.2	96.2
EF vs EE	**90.0**	**90.0**	85.0	**94.4**	**94.4**	85.0	**85.0**	**85.0**	85.0	**89.5**	**89.5**	85.0	**98.1**	97.0	95.5
YY vs EE	**97.5**	90.0	**97.5**	95.2	94.4	**100.**	**100.**	85.0	95.0	**97.6**	89.5	97.4	99.5	98.8	**100.**
EY vs EE	**92.5**	**92.5**	87.5	**94.7**	**94.7**	89.5	**90.0**	**90.0**	85.0	**92.3**	**92.3**	87.2	**99.5**	98.4	98.2

Table 5.2: Logistic Regression scores evaluated through LOO Crossvalidation for age-groups Binary Classification.

	Accuracy			Precision			Recall			F1			ROC-AUC		
	Text	List	Text+List	Text	List	Text+List	Text	List	Text+List	Text	List	Text+List	Text	List	Text+List
YY vs EY	57.5	**65.0**	**65.0**	58.8	**68.8**	**68.8**	50.0	**55.0**	**55.0**	54.1	**61.1**	**61.1**	56.5	**68.2**	68.0
EY vs EF	**67.5**	65.0	62.5	**68.4**	62.5	61.9	65.0	**75.0**	65.0	66.7	**68.2**	63.4	72.5	65.8	**73.5**
EF vs EE	72.5	**77.5**	72.5	69.6	**78.9**	71.4	**80.0**	75.0	75.0	74.4	**76.9**	73.2	80.5	81.2	**83.8**
YY vs EE	90.0	**92.5**	**92.5**	94.4	94.7	**100.**	85.0	**90.0**	85.0	89.5	**92.3**	91.9	93.0	94.2	**98.2**
EY vs EE	85.0	**87.5**	85.0	88.9	**89.5**	85.0	80.0	**85.0**	**85.0**	84.2	**87.2**	85.0	88.8	**95.8**	95.0

cording to the average contribution in features coalitions, i.e. its Shapley value. It is then possible to explore the role of each feature in the classification of single samples, independently from the fact that they have been learned by the model during the training step. The sample prediction represents the sum of the feature contribution starting from the baseline: if a feature has a positive impact, it influences the prediction in favour of class 1 and vice versa. This step was useful to understand, for each sample and each age-group, how much every indicator leads the model to predict class 0 or 1.

5.1.6 Results

The performance metrics for each classification task and dataset are reported in Tab. 6.4, with the Catboost classifier, and in Tab. 5.2, with the Logistic Regression. As expected, the Catboost algorithm achieved the higher performance. In the tables, for each metric, the higher scores among datasets are written in bold numbers.

The detailed outcomes of 3 classification tasks are reported in Fig. 6.6: EYvsEF in the first column of the figure, EFvsEE in the second column and EYvsEE in the third. The first and the second tasks involved the most interesting class for monitoring purpose (the EF, with individuals in the range 60-69 years of age) and its closest two classes in terms of age ranges (40-59 and 70+ respectively). The third task was instead aimed at evaluating how much the age gap 60-to-69 years of age improved the binary classification between the younger and the older groups of individuals. The row (a) shows the ROC-AUC performance obtained with Catboost trained and evaluated over Text, List and Text+List datasets with the LOO strategy. In all the cases, the results from Text datasets achieved the higher ROC-AUC. For these reasons, the plots in rows (b, c, d) showed the results from

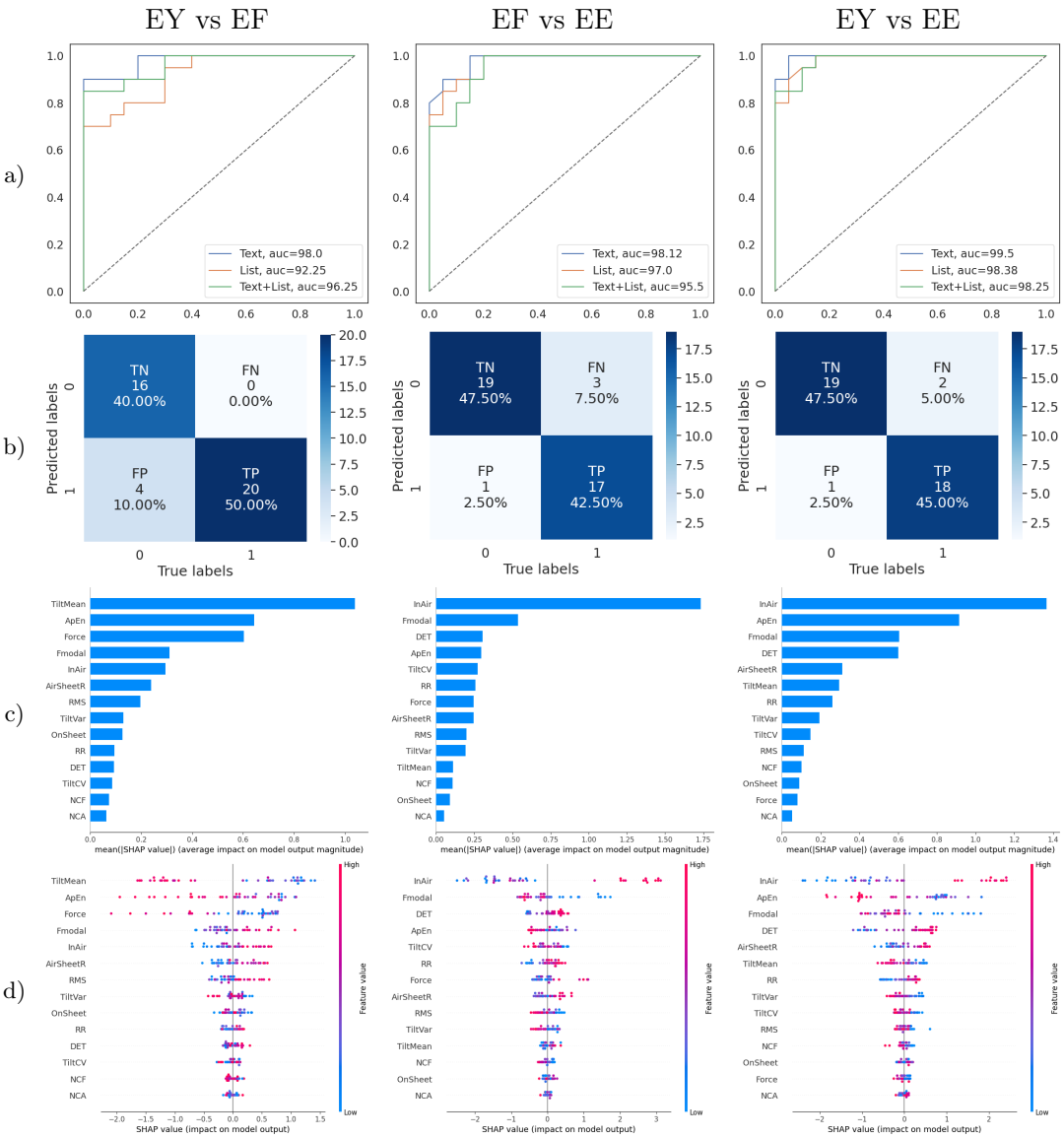

Figure 5.2: Classification performances and model explanation plots for the tasks EY vs EF, EF vs EE and EY vs EE: the ROC-AUC metrics achieved by Text, List and Text+List indicators are in row a); the confusion matrices are row b); rows c) and d) reports the absolute average shapely values and the shapely value of the features for each sample respectively.

Text datasets only. The row (b) shows the confusion matrices and the rows (c, d) show the final models SHAP feature rankings trained on the whole task datasets and tuned via LOO. While row (c) displays the absolute impact of the features, row(d) shows the same ranking and explains how the samples were predicted according to their features values. Each dot in the figures in d) represent the shapely value of the indicator for a particular sample. The blue-red colour scale indicates the indicator value (low to high) and the negative shapely values pushed the prediction towards class 0 (the younger group), while the positive values favoured the classification of the subject in class 1 (the older group). The outcomes of this three tasks are detailed in the following subsections.

EY vs EF

In EY vs EF task, the ROC curves revealed that the Catboost models trained using the Text data achieved the best results with an AUC of 98.0%. In the associated confusion matrix referred to Text data, the absence of False Negatives is translated in 100% Recall, while overall, there are 4 False Positive miss-classifications. According to the Shapley values of the final model, this task was strongly affected by the mean tilt angle, the approximate entropy and the writing force indicators. Indeed, higher mean tilt values were related to younger subjects belonging to age-class EY like for approximate entropy and writing force.

EF vs EE

Finally, in EE vs EF task, the ROC curves revealed that the Catboost models trained using the Text data achieved the best results with an AUC of 98.12%, which means it is possible to discriminate also between the two oldest classes. In the associated confusion matrix referred to text data, there is almost the same behaviour of the previous task, having only 1 False Positive and 3 False Negative predictions translated into 90% of Accuracy and 94.4% of Precision. According to the Shapley values of the final model, the t_{air} indicator turned out to be the core feature for this binary classification. The model strongly related high values of t_{air} indicator to the older class EE.

EY vs EE

In EY vs EE task, the ROC curves revealed that the Catboost models trained using the Text data achieved the best results with an AUC of 99.5%, i.e. the predictions of unseen data are almost perfect. In the associated confusion matrix referred to text data, there are only 1 False Positive and 2 False Negatives, which lead to a balanced F1 score equal to 92.3%. According to the Shapley values of the final model, this task was strongly affected by t_{air} and $ApEn$ indicators and more softly by F_{modal} and DET ones. The

	Resting tremor	Hand tremor	Rigidity	Digit dexterity	Hand movements	Rapid alternate hand movements	Bradikinesia
Mean score (0-4)	0.85±0.9	0.19±0.4	1.62±0.8	1.73±0.82	1.7±0.8	1.6±0.7	1.73±0.78

Table 5.3: Average scores for the tremor measurements of the PD patients

t_{air} and DET higher values pushed predictions favouring the older class EE, while high $ApEn$ and F_{modal} values were more linked to younger subjects belonging to class EY.

5.2 Test on PD subjects

The ability of the handwriting indicators in discriminating healthy or pathological writing gesture has been investigated by performing a classification task between control age-matched older adults and PD patients.

5.2.1 Methods

Free text handwriting data were collected from 20 PD patients (10 males, 73.6±6.6 y.o.a.; 10 females, 73.4±8.14 y.o.a.) in the IRCCS Fondazione Maugeri (Milan), using the acquisition protocol described in Sec. 5.1.1. Inclusion criteria were a maximum hand tremor score (0-4) of 1, a maximum digit dexterity score (0-4) of 3 and a maximum hand movement score (0-4) of 3. The average values (from 0 to 4) of tremor measurement for the PD patients are listed in Tab. 5.3 For the healthy control, 10 subject were considered from the grop EF and 10 from grop EE (see Sec. 5.1.1), all aged over 65.

The set of indicators described in Sec. 5.1.2 were calculated for each subject and included in the classification dataset. In total, 40 samples with 14 features were obtained. The binary classification problem with the two classes (*healthy* and *PD*) was set. The task was performed using the Catboost classifier with a weighted *logloss* cost function, implemented in the Python library Scikit-Learn [176]. Accuracy, Precision and Recall were considered as performance metrics, estimated using the leave-one-out crossvalidation (LooCV). At last, the model explanation method SHAP, described in Sec. 5.1.5, was applied to have a better insights on the contribution of each indicator in the classification.

This classification setting could be interpreted in two different ways. The more conventional one, the *screening* task, was intended to spot all the PD patients using the classifier as a test to evaluate the presence of the condition.

In this case, having a false negative would have been a more serious error than having a false positive, as further tests may confirm the illness.

The other interpretation instead is referred to the handwriting monitoring, with the objective of detecting signs of non-physiological decline. Here, one assumes that the individual to be monitored is in a known particular health status, which can be healthy or at the early stages of the PD. Now, the classifier could be used to check, at certain points in time, whether the subject's status has turned in a particular stage that might be associated with some condition of PD. In this second case, one is interested in avoiding, as much as possible, the false positive. As a possible alarm bell would result oversight.

The balance between Sensitivity and Recall can be adjusted by tuning the threshold as indicated by the ROC analysis. However, weights could also been set to increase (or decrease) the importance of the two classes during the training phase of the classifier. In this way, the model is pushed to favour one class over the other in the learning phase. Class weights were chose to achieve different performance while training the classifier.

5.2.2 Results

Balanced weights classes

The results of the binary classification using the same weights are reported. The confusion matrix is shown in Fig. 5.3 panel (a). and the outcomes were an Accuracy score of 92%, a score of 95% for Precision and 90% for Recall. Figure 5.3 panel (b) shows the ROC curve for the healthy-PD classification task, with balanced weights.

Higher weight to the PD class

In the training process, the PD class was weighted the 2.0% more[4]. The confusion matrix is shown in Fig. 5.4 (a) and the outcomes were an Accuracy score of 95%, a score of 100% for Recall and a Precision score of 90%. Note that, in this case, the Recall has become maximal.

Higher weight to the healthy class

In the training process, the Healthy class was weighted the 0.5% more. The confusion matrix is shown in Fig. 5.4 (b) and the outcomes were an Accuracy score of 92%, a Recall score 85% of and a Precision score of 100%. In this case instead, the Precision resulted maximal.

[4]The weight increments was set empirically by choosing the best results

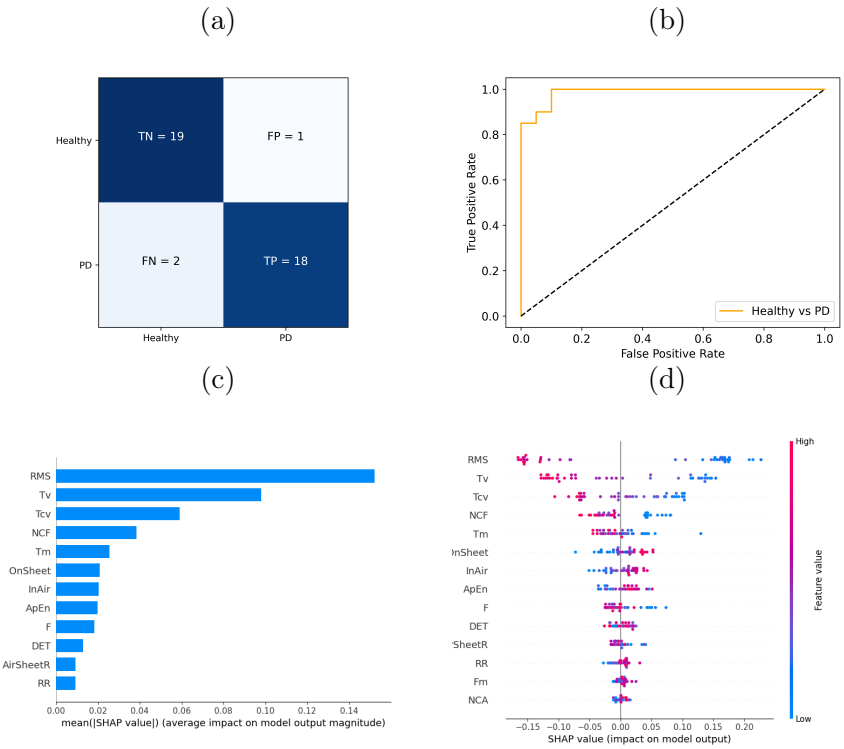

Figure 5.3: Classification performance for healthy-PD with balanced class weights and model explanation. Confusion matrix in panel (a), ROC in panel (b), average Shapely value in panel (c) and samples Shapely value in panel (d).

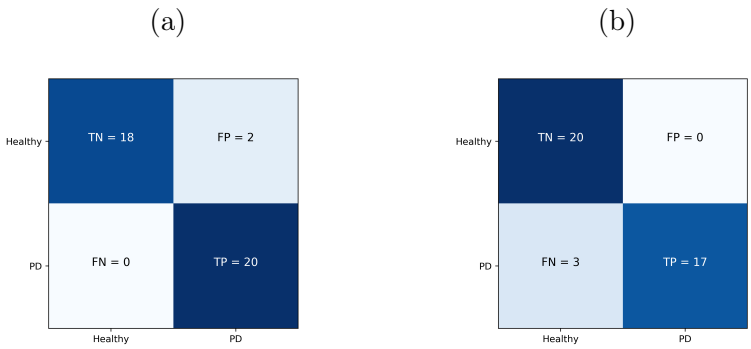

Figure 5.4: Confusion matrices for the classification healthy-PD with unbalanced weights, + 2.0% of weight for the PD class (panel a) and + 0.4% of weight for the healthy class (panel b).

Model explanation

The plots in Fig. 5.3 panel (c) and (d) represent the outputs of the application of the model explanation technique SHAP to the classification model trained for the healthy-PD classification with balanced weights class. The shapely values were computed for each features and they indicated their singular impact to the classification outcome of the samples. The bar plot in panel (c) shows the average absolute shapely value among all the samples. It could be interpreted as the average impact the feature had in the classification task, therefore it indicated the feature relevance rank. The plot on the left, in panel (d), shows a scatter representation of the shapely values of all the samples. The dots are displaced in rows, one for feature, and the x-axis reports the shapely value. The impact of a feature was in favour of the healthy class if its shapely value was negative, while it pushed the outcome towards the PD class if its shapely value was positive. The colour-map (blue for the lower values, red for the higher values) indicated the feature magnitude (in relation to its measurement unit).

5.3 Discussion

Age-Groups Classification

In this work, the utility of the quantitative analysis of handwriting and tremor was studied in the discrimination between healthy subjects belonging to different age groups. A novel ink pen for the handwriting data acquisition was used in paper-pen tasks, mimicking daily writing. Indeed, participants were asked to write a short free text and a grocery list without constraints on the content or writing modality. This particular setting was chosen to maximise the study ecological validity, with the ultimate goal of leveraging these findings for home-based solutions dedicated at the early detection of decline in seniors. Therefore, particular attention is paid to the correct classification of the individual's age, since the association of its handwriting features with an older age group could be interpreted as a clinically relevant anomaly [177].

One of the last and most performing ML classification algorithm, Catboost, and a more traditional one, Logistic Regression, were used to classify the subjects in one of the four age-groups, starting from the set of handwriting indicators computed from the raw free-text and grocery list data. Results showed that the Catboost algorithm outperformed the logistic regression in almost all the tasks and the datasets considered in this study. The improvements were sensibly higher in the classifications between groups with close age ranges (the first pool of tasks), where the differences in the individuals' handwriting were expected to be minimal. That confirmed the superior sensibility of Catboost to the changes in the handwriting indicators

with respect to a baseline estimator.

In the first pool of tasks, the classifications between the groups with individuals in close age ranges (i.e. *YY-EY*, *EY-EF* and *EF-EE*) were considered. The objective of these tasks was that to test the sensibility of the models to small variations in the handwriting decline, which were expected to find between healthy individuals with limited age differences [16]. Very good to excellent performance (Accuracy between 82.5% and 90%, Precision from 81.8% to 94.4%, Recall from 75% to 100% and ROC-AUC 92.2% to 99.5%) were obtained in the classification of the first pool, considering all the three datasets composed by the indicators computed form the Text, List and joined Text-List data. These scores revealed the good ability of the models in detecting slight handwriting variations in healthy subjects. Therefore, an high sensibility to the changes in the handwriting data due to an abnormal or pathological ageing decline may be expected [16]. In the second pool of tasks instead, the classifications between more distant age groups was considered. As expected, outcomes were commonly higher in these tasks because more evident differences should have been visible in handwriting. In the classification between *YY-EE*, i.e. the more distanced classes, the best Accuracy (97.5%) was achieved using the only text indicators, and the joined set of the Text and List indicators. Perfect Precision and ROC-AUC (100%) were obtained using the Text+Tist data and, a perfect Recall using the sole Text data. The last setting, *EY*vs*EE*, showed higher evaluation metrics with both Text and List data, all over 92.3%.

For the classification between the younger groups, *YY* and *EY*, i.e. 18-40 and 41-60 years of age, the List turned out to be the data with which the model achieved better outcomes of Accuracy, Recall, F1 and ROC-AUC. In the other adjacent groups tasks, *EY-EF* and *EF-EE* (i.e. 41-60 vs 61-70 years of age and 61-70 vs 71+ years of age) the best performance were obtained with the Text, in almost all the cases. This minor task-dependency of the classification outcomes might be related to some differences between the Text and the List tests. At first, Text samples were longer, as they contained many words. Then, the writing dynamics might result partly affected by the type of task since every element of the List was written in a new line and short words, as articles and propositions, were less frequent. Besides, writing a free text generally required a more significant cognitive effort by the subjects. However, the differences in the results were little and, given the relatively small amount of samples, random factors could not be excluded. Further investigations are envisaged. Results suggest that both the data acquisition modalities are still valid and contain intrinsic age-related information.

In this work, the experiments were deeper analysed using the model explanation technique SHAP. It was useful to understand the impact of each handwriting indicator in the different tasks and to see their behaviour. In

this study, the outcomes and the analysis of three particular classifications are detailed: the first two belonged to the first pool and included the class of the individuals in the range 61-70 years (EF), which is the more critical for the early detection of decline purpose; the third consisted of the classification between the individuals in the age ranges 41-60 and 70+, and it was aimed at showing the more strongly marked difference in the handwriting characteristic of the two classes.

The first classification, the groups aged 41 to 60 and 61 to 70 (EY and EF) were considered. The two groups represented respectively a population of healthy subject in which the effects of the age-decline is absent and a population in which a degradation in the physical or cognitive functionality might be at an initial stage [17, 18]. As shown in Fig. 6.6 row b, the handwriting indicators were able to correctly classify all the individuals in the range 61-70 years of age (with a recall score of 100%), while four subjects in the range 41-60 were miss-classified (precision score equal to 84.2%). The outcomes confirmed previous findings in literature in which handwriting has been observed to undergo perceptible variations in middle younger and older adults [178]. According to Walton [8], handwriting characteristics may be stable at least within 5 years in healthy subjects. Indeed, the four false positive subject were all older than 52 years. Two of them were aged 65+. For this reason, it was likely that their writing characteristics resulted closer to the older group. The model explanation (Fig. 6.6, rows c and d) revealed both handwriting dynamics and tremor features among the more influential ones. The inclination of the pen ($Tilt_{mean}$) resulted the most important feature in the EY vs EF classification. In accordance with Marzinotto et al. [178], an higher inclination of pen (on the right) was typical in the mid-older adults (EF). The approximate entropy ($ApEn$) also played a major role, indicating a lower predictability of the younger class handwriting time series. This result corresponded with the findings of previous works [160] where, using similar experimental settings, significant differences were found between age groups. The trend which saw entropy decreasing with age was in line with previous literature in the study of resting and postural tremor in younger and older adults [20, 15, 153]. Even if its variations did not resulted statistically significant among differently aged groups [160], the writing force (F) emerged as the third most meaningful feature in the EY-EF classification. The predictions were moved towards the older group (EF) when the force values were lower. That was in accordance the study by Engel Yeger et al. [9] in 2012, Caligiuri [164] in 2014 and Marzinotto et al. [178] in 2016. In the following four features, sorted by decreasing importance, two frequency-domain and two temporal parameters were found. The modal frequency had no significance in the statistical group differences in the previous study (see Section 4.4.3) [160], however it impacted the classification EY vs EF, as higher values were associated to the older class. The same behaviour resulted for RMS. Preceding studies showed that some neurological con-

ditions, such as PD, could affect the modal frequency [151], while no clear age effect on this parameter have been showed. The impact of temporal indicators (t_{air} and $t_{a/s}$) was considerable and confirmed the tendency of the older class to have longer non-writing moments, found in the previous work (in Section 4.4.3) [160] and others [129, 179]. The rest of the indicators exhibited a lower impact in the classification between EY and EF.

The second classification, between the groups aged 61 to 70 and 70+ (EF and EE), was the most relevant to study the suitability of the our approach in the early detection of decline scenario. In a normal ageing process, a physical or cognitive decline is expected to be more consistent in the older group, of the people aged 70+ [17, 18]. Therefore, whenever an individual in the younger group (aged in the range 61-70) is associated to the older one, it might be interpret as a sign of abnormal decline. In this task, the handwriting indicators were used to discriminate individuals aged 61 to 70 from those aged 71+ with high scores of performance. Our results showed that the EF-EE classifier may be suitable for the decline monitoring application because of its high Precision of 94.4%. Only 1 subject to 20 was wrongly classified as older, while the false negative were 3 (Fig. 6.6, row b). The model explanation (Fig. 6.6, rows c and d) revealed that the in-air time parameter (t_{air}) was way more influential in the classification than all the others. As for the other tasks, higher t_{air} were associate with individuals of the older class. Modal frequency was the second indicator for importance, whit an impact lower that the 31.4% w.r.t the first feature. Quite similar impact had the other indicators, with the frequency and non-linear features in higher ranking positions. The pen inclination ($Tilt_{mean}$) ended among the last important indicators, although its variation ($Tilt_{cv}$) resulted having a more considerable impact. Nonetheless, all the indicators maintained the same behaviour of the previous tasks, so this corroborated the consistency of the handwriting measurement variations with age.

The third classification was between the EY and the EE groups, with individuals in the ranges 41-60 and 70+ years of age. The level of decline was expected to be very different among the healthy subjects' populations included in this tasks. As a consequence, the ability of the model in discriminating between these classes of individuals using the handwriting indicators resulted increased. The Accuracy score was equal to 92.5% and the Precision was notably higher, with 94.7%, at the expense of a minor Recall, equal to 90%, w.r.t the previous task. In fact only one subject in EF and two in EE were wrongly classified. The model explanation (Fig. 6.6, rows c and d) showed almost the same indicators among the more meaningful, however some important differences appeared. The $Tilt_{mean}$ dropped from the first to the sixth position in the impact ranking. Still keeping the same behaviour. In this task, t_{air} emerged to have the higher impact, with the same trend of showing higher values in the older groups. The writing force dropped from the third to the penultimate position, while determinism DET

raised up to the fourth place, with the same impact of f_{modal}. Determinism was likely to increase with age as the influence of the predictable tremor components became more persistent. All the handwriting indicators showed the same behaviour in the classifications of EY with EF and EE. Major changes were found in the impact level of the features associated to tremor, which resulted more determinant in the discrimination between the two more distant groups, EY and EE.

The model explanation revealed that the impact of the handwriting indicators was task dependent, i.e. it changes according to the age ranges considered in the classifications. These differences in the feature importance highlighted the complexity of the age-driven decline in handwriting as the sensitivity of some indicators showed age-dependency. However, the behaviour of the indicators in the different age intervals was analogous to the previous findings in literature in populations of healthy subjects. This result reinforced the interpretation of the models, giving the possibility to understand their decisions as they relied to known handwriting-related quantities.

Healthy-Pathological Classification

The classification between healthy subject and PD patients showed very good results, with an accuracy score equal or higher than 92% in all the cases here considered. For the screening purpose, the model was invited to pay more attention to the healthy class, by increasing its weight. It resulted in a maximum Precision of 100%. On the other side, increasing the weight of the PD class, the Recall reached the 100%. This made the classifier more suitable for screening purpose.

As predictable, the model explanation revealed that the tremor related indicators had the larger impact in the classification between healthy subjects and PD patients. As reported in the study by O'Suilleabhain et al. [151], the frequency domain indicators (RMS and modal frequency) appeared to be affected by neurological conditions such as PD. In the presented classification, higher values of RMS and modal frequency were associated to the PD group. The second and the third features, ranked by importance, were the approximate entropy and the determinism respectively. The entropy measure showed a decrease in the PD group, suggesting the tendency of developing more regular tremor patterns in individuals affected by the Parkinson's. Contrariwise, the determinism has been shown to increase as the entropy decreases. Thus confirming an increase of predictability in the handwriting dynamics in the PD individuals. With the same behaviour, also the recurrence ratio resulted to increase in the PD group. More regular tremor patterns has been found indeed in previous studies in PD patients [153].

The writing force, in the sixth position in the importance ranking, resulted higher for the PD patients. This finding was in accordance with the

study by Walton, in 1997, [8], which observed and on/off-like pressure activation in PD. In that study, the writing force applied by the PD participants was uniformly moderate, but suddenly, for many word, it become very heavy.

At last, the temporal features had a minor impact in the classification. However their behaviour suggested that the PDs spent longer time intervals with the pen lifted from the paper sheet. The same study by Walton [8] reported that the PDs raised their pens within words at least twice as often as controls did. This confirm a lower fluency in handwriting for PDs, also attributable to the slower hand movements.

Chapter 6

Detection of Cognitive Decline

This chapter focuses on the cognitive decline in elders. In medicine, any degradation in the mental functionality that is sufficient to compromise independent life and the execution of common activities in the individual is referred as dementia. It is indeed a neurodegenerative syndrome that may be caused by a great variety of conditions, including primary neurologic, neuropsychiatric and medical conditions [180]. It is also very common that the cause of a dementia syndrome in patients consists of multiple diseases [30]. The global prevalence of dementia reaches the 7% of people aged 65+. However in developed countries, values are up to 8%-10% due to the increased life spans [181]. Neurodegenerative syndromes of dementia, such as Alzheimer's disease (AD) and dementia with Lewy bodies, are the most common in older adults. In fact, the advancing age is among the major risks of cognitive degradation [180].

Alzheimer's disease has higher prevalence (up to 30%) in elderly aged over 85. Nevertheless a prevalence between 5% and 6% is estimated in all individuals aged 65+. In the 5% of the AD patients, the disease occur before the age of 65, in a type of disease known as *early onset AD* (EOAD). AD with onset after age 65 is called *late onset AD* (LOAD). In the early stages of AD, individuals assist to a slowly progressive memory decline and, in more rare cases, also behavioural or language symptoms are visible. As depicted in Fig. 6.1, both EOAD and LOAD are characterised by a preclinical phase that may begin 20 years earlier. In this condition, individuals can show occasionally forgetfulness, irritability and a low mood [182]. As the cognitive degradation progresses, the individual enter a condition called mild cognitive impairments (MCI), where some symptoms can occur before any conspicuous functional decline has manifested [5]. The detection of MCI can be further complicated because symptoms may be confused with those of normal ageing [16]. In the later stage, the medial temporal lobe atrophies, including other surrounding structures as the hippocampi. Brain pathology and dysfunction can be found by looking for plaque deposition in multiple

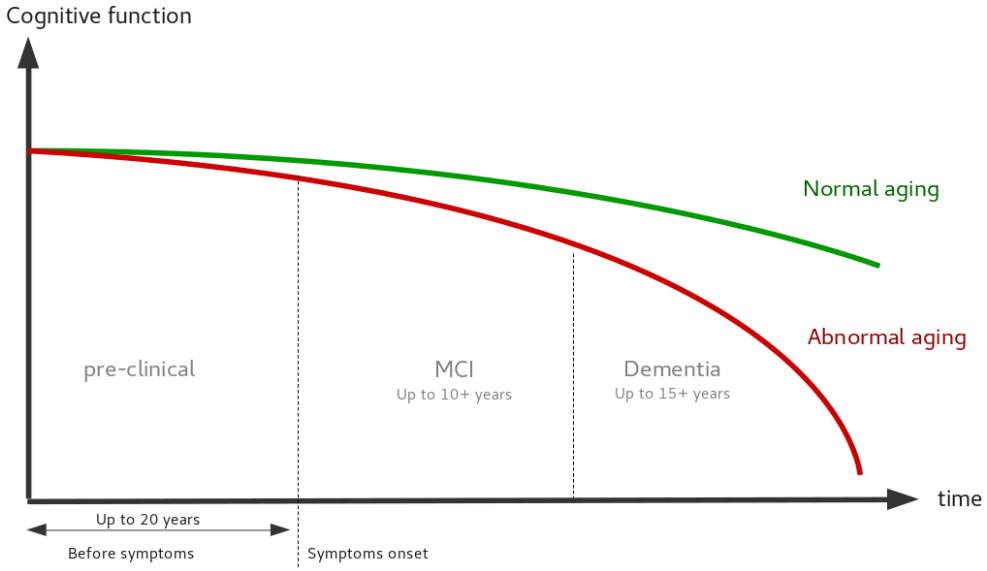

Figure 6.1: The continuum of cognitive decline.

regions of the brain using amyloid positron emission tomography (PET) and bilateral temporoparietal hypometabolism by fluorodeoxyglucose-PET, respectively. [183].

Currently, pharmacologic treatments which modifies the AD are not available. Therefore, the clinical and translational research focuses on the early detection of cognitive decline, in addition to the therapeutic targeting of the disease's histopathology [184].

In the clinical practice, the assessment of dementia includes three elements: i) the investigation of the clinical history; ii) the neurological examination (with the assessment of the mental status) and iii) selected instrumental and laboratory exams [5]. Very important for the detection of the cognitive decline, also for reasons related to the non-invasiveness, is a detailed mental status examination. It analyses various domains of mental function, among which basic attention, memory, visuospatial abilities and executive functions.

A screening tool for the dementia severity is the mini-mental state examination (MMSE) [185]. However, since its sensitivity can be low for some elder populations, a battery of several verbal or paper-based neuropsychological tests are further carried out to deepen the investigation on the symptoms and to achieve a more precise diagnosis. Each test is evaluated with the assignment of a numerical score according standardised rules. The outcome of the cognitive assessment is based on the comparison of the scores the subjects received with reference values [186]. The scoring of some tests, such as the two variants of the trial making test (TMT) [187], is pretty straightforward

(i.e. its calculation is just based on the time of execution of the test variants). For other tests, as the Rey Osterrieth complex figure (ROCF) test, the evaluation can be more convoluted [188], since they account for various cognitive functions.

Digital version of simpler tests, like the TMT, has been also proposed in the eHealth scenario to enable their (remote) self-administration in community dwelling elders [25]. In the context of home monitoring, these type of technologies might improve the effectiveness of early detection of cognitive decline as the clinician can timely receive the negative scores as soon as the individual performed the test at home. For the ROCF test digital implementation instead, major limitations are represented by the difficulty in developing an accurate reliable automatic scoring system and the high inter rater scoring variability. In addition, the sole remote inspection of the test output by the clinician does not supply sufficient material for the evaluation.

In this study, a computer algorithm is developed to give assistance to the clinician in the evaluation of the ROCF test, even when just the output image of the ROCF copy is available. The algorithm consisted in a decision support system (DSS) which help the evaluation of the test using knowledge of previous administered tests.

The rest of this chapter is organised as follows: Sec. 6.1 induces and explains the ROCF test and its evaluation method; Sec. 6.2 presents the data collection and pre-processing for the implementation of the DSS, the methods for the the ROCF-copy patterns evaluation and the diagnosis formulation algorithms; then, results are shown in Sec. 6.3 and discussed in Sec. 6.4.

6.1 The Rey-Osterrieth Complex Figure Test

The Rey-Osterrieth Complex Figure (ROCF) [189], in Fig. 6.2, is a nonverbal neuropsychological test widely used to evaluate visuo-constructional abilities and nonverbal memory [190], also in presence of motor symptoms [191], in children, adults and older adults. Neuropsychological testing is the main non-invasive diagnostic instrument for the assessment of impaired conditions, such as mild cognitive impairment (MCI) and dementia, in elder individuals [192]. The mental assessment consists of a battery of several tests whose score give detailed information about various cognitive domains (visuospatial function, memory, attention, executive function and language). It generally involves two variants of the ROCF test: the copy and the recall. In the copy test, the subject have to sketch the ROCF while looking at a template figure as reference. In the recall test instead, the task is repeated after 30 minutes without the reference image.

The ROCF is articulated in 18 geometrical patterns (see Fig 6.5). The outcome of the test consists of a total score computed as the sum of the

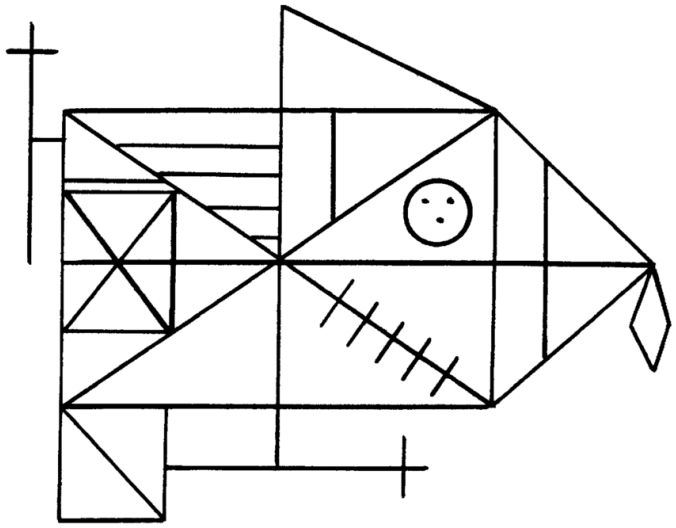

Figure 6.2: The Rey-Osterrieth complex figure (ROCF).

single scores the clinician assigned to each pattern drawn by the subject. A standard protocol for the assessment of the ROCF test is the Osterrieth system [26]. A numerical score (between 0 and 2) is given to each pattern according to the quality of its representation: 0 if indistinguishable or absent, 0.5 if deformed and misplaced, 1 if correct and misplaced or just deformed, and 2 if correct and well placed in the figure. The higher the total score, the better the cognitive performance of the subject are. According to Bertolani et al. [193], an age and schooling-corrected total score in the copy test lower than 28 indicates a deficit, while a score between 28 and 32 may suggest cognitive decline.

Several studies however investigated the reliability of the ROCF test scoring and, some of them, reported poor rater agreement [27, 194]. Systematic differences were found on both the copy and recall variants, as statistically significant mean differences were observed, with a possibly of score variations approaching the 20% [27]. With these inter-rater discrepancies, it seems plausible that clinical assessment can be affected by an aberrant score. However, the impact on the determination of the diagnosis in the clinical examination can be little, since it does not depend on the sole ROCF test. On the contrary, the investigation of the relationship between central nervous system impairments with changes in memory functions and the study of the subject's cognitive variation over time could be altered by minor differences in the ROCF test scoring [28, 195].

Some works developed computer algorithms to automate the ROCF scoring with the aim of improving the inter-rater reliability and the sensitivity of the test [196, 197]. Although high performance in accuracy were obtained

using multiple deep neural networks, the correlations between the automated scores and the human raters' scores were not strictly equivalent [197].

In this study, a decision support system (DSS) for the evaluation of the ROCF copy test was developed with the goal of supplying qualitative information about the judgement of each pattern and of suggesting the most probable diagnosis for the subject among normal, MCI and dementia. The system was not aimed at replicating and replacing the human scoring, but it was intended to be an additional tool for the experts in the subject's examination process. Indeed, the design was focused to provide an instrument which helps improve the accuracy of the diagnosis, leveraging on the knowledge about numerous previous assessments of the ROCF-copy test and methods from explainable artificial intelligence (AI). The DSS required the initial human action of selecting strategic point in the image. Then it run computer vision and deep learning (DL) algorithms to detect and evaluate the patterns. The software returned 18 categorical scores, one per pattern, and the most probable diagnosis for the subject. The DSS algorithms were calibrated and trained using retrospective copy-test records from 250 mid-aged to older individuals (healthy, MCI and with dementia). The system performance were assessed by computing the accuracy in the pattern evaluation and various classification metrics in the formulation of the diagnosis.

6.2 Design of the Decision Support System

6.2.1 Participants and Data Collection

For this study retrospective data from subjects and neurological patients who underwent a neuro-psychological examination in the Istituto Palazzolo, Fondazione Don Carlo Gnocchi in Milan (Italy), were acquired within a time period from Jan 2017 to Dec 2018. Digital versions of the ROCF copies were collected by scanning the paper records filed in the annual clinical registers. A total number of 57 samples were acquired for the normal subjects, 131 for the MCI and 62 for the patients affected by dementia. The figures were scanned with a resolution of 300dpi and saved as portable graphics format (PNG) in RGB colours. Personal data (age, years of education) and other information such as the date of the visit and the MMSE score were also retained [1].

6.2.2 Categorical Patterns Evaluation

Each ROCF sample in our dataset was associated to an individual labelled as *normal*, *MCI* or *dementia*, according to the outcome of his neuropsycho-

[1]The study was approved by the internal review board of the Instituto Palazzolo.

Table 6.1: The count of the patterns, divided for each label

Pattern number	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*
Omitted	15	26	55	170	43	67	33	34	70	55	79	36	73	18	24	39	34	34
Distorted	104	139	76	16	107	0	0	153	30	55	23	101	65	55	46	11	94	112
Misplaced	42	0	43	0	8	12	56	0	36	10	8	0	8	15	24	38	0	0
Correct	89	85	76	64	92	171	161	63	114	130	140	113	104	162	156	162	122	104

logical evaluation. The eighteen patterns of the ROCF image were manually inspected and scored using expert-based rules. A 4-elements categorical scoring system was chosen, rather the standard Osterrieth scoring, to simplify the automatic pattern evaluation of the DSS. Each pattern was independently examined and one of the following scores was assigned when this specific conditions applied[2]:

- 0 *(omitted)*, if the pattern was not represented nor recognisable in the figure;
- 1 *(distorted)*, if a distorted[3], yet recognisable, version of the template pattern was represented in the figure;
- 2 *(misplaced)* if the pattern was not distorted but placed differently from the expected location;
- 3 *(correct)* if all the previous conditions did not applied.

A 250 rows-dataset was build by inserting the 18 human-assigned scores to each pattern as columns for each image. The last column reported the individual's clinical outcome. The count of the patterns, divided for each label, is reported in Table 6.1.

6.2.3 Image Pre-processing

The preparatory phase of the image analysis consisted in the noise removal, which was mostly characterised by the clinician's annotations during the ROCF copy evaluation. The clinician's signs consisted in the numerical score they sketched upon the drawing to keep memory about the already evaluated patterns. In all the images collected, scores were written using red and green ink pen, while the ROCF lines were drawn with a common pencil. This marked difference between the clinician's and patient's sign allowed the noise removal through the use of colour filtering techniques.

The images were converted in the hue-saturation-value (HSV) colour space [198] to identify the colour shades to remove more easily. In the HSV format, the colour of each pixel in the image is described by three values,

[2]The labels were treated as categorical. Yet they ranked from 0 to 3 according to the correctness of the drawing

[3]The patterns which were not topological equivalent to the template were labelled as *distorted*.

in the range 0°-360°, 0-100 and 0-100 respectively, which defined the coordinates of a cylindrical geometry. The first number represents the hue spectrum where 0° is the red primary, 120° is the green primary and 140° is the blue primary; the second coordinate indicates the saturation level as the radial distance from the center and the last number expresses the mixture value form dark to light, as the height of the cylinder. The HSV triplets [0 30 10] was found, for red ink, and [160 30 10], for green ink, as the optimal values to identify the noisy pixels in the image. Then, the colour of those pixels was replaced with the most frequent pixel value, which was assumed to be the background colour of the paper sheet. Other spurious noise component were attenuated using low pass and median filtering with 3x3 kernels.

After the noise removal, the image were binarized. Unimodal thresholding [199] was applied to separate the pixels belonging to the drawing (set to 0 grey-scale intensity, *black*) from the ones belonging to the page (set to 255 in grey-scale, *white*). The handmade drawings however were imprecise. That resulted in strokes with irregular thickness and non-perfectly closed shapes in the binarized images. To attenuate the effect of these additive disturbances, image erosion [200] with a 9x9 kernel was used to cover all the gaps between close tracts. Then, a Skeletonization algorithm [201] was applied to the negative-binary image to obtain one pixel wide lines. At last, the drawing lines were dilated with a 3x3 kernel to enhance the objects in the image.

All the images resulted very different in shapes, dimensions and proportion. Therefore a figure standardisation was required before performing further analysis. The user od the DSS was asked to manually select five reference point in the image. Those points, indicated in Fig. 6.3(a), identified the four vertices of the main rectangle in the ROCF plus the rightmost point of the figure, which coincided with a vertex of the right triangle. The reference points were used to perform an image homography [202] to match the reference points to their respective locations in the template model, preserving the structure of the patient's drawing. A binary representation of the original ROCF centred in a 428x733 pixels biding box was retained as template model image. An example of the figure standardisation is shown in Fig. 6.3 (b). Noise removal, image processing and the homography were implemented using the image processing Python libraries OpenCV 4.5.1.48 [203].

6.2.4 Workflow of the DSS

After the image preprocessing, the analysis of a sample ROCF image was implemented in two main stages: the *pattern evaluation* and the *diagnosis formulation*. Pattern evaluation in turn consisted of an initial *detection* step, in which a pattern was searched in the figure to determine its presence or absence, and an *evaluation* step in which a label was assigned to the pattern.

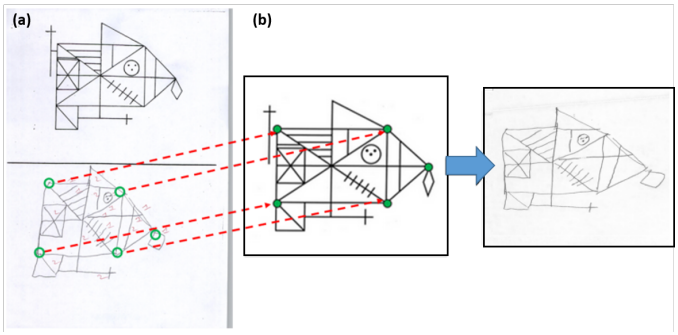

Figure 6.3: Example of the standardisation of a ROCF sample, Panel (a) represent the scan the of paper sheet where the template (in the upper part) and the patient's replica (lower part) of the ROCF are figured. The green dots indicates the reference points selected by the examiner and the red lines associate the drawing reference points to the ones of the template model. Panel (b) shows the ROCF sample after the isomorphyc transformation.

A schematic representation of the DSS workflow is reported in Fig. 6.4. For their evaluation, the patterns were addressed as *simple* or *complex* and computer vision (CV) or deep leaning (DL) methods were applied in the former and in the latter case respectively. The patterns were sorted into the two categories, as showed in Fig. 6.5, according to their shape and elements. The initial coordinates of the area in the image in which a pattern was searched, i.e. the region of interest (ROI), was chosen from the template model as shown by the red shapes in Fig. 6.5. The initial ROI of a pattern (noted with ROI_0^p, where p indicates the pattern) could be adjusted during the analysis, using the information of those previously detected. For the detection of some patterns, additional rotated variants of the initial ROI

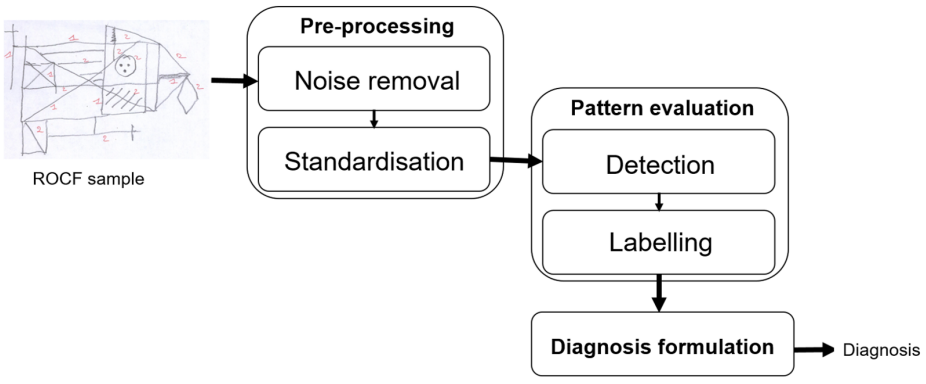

Figure 6.4: The workflow of the DSS.

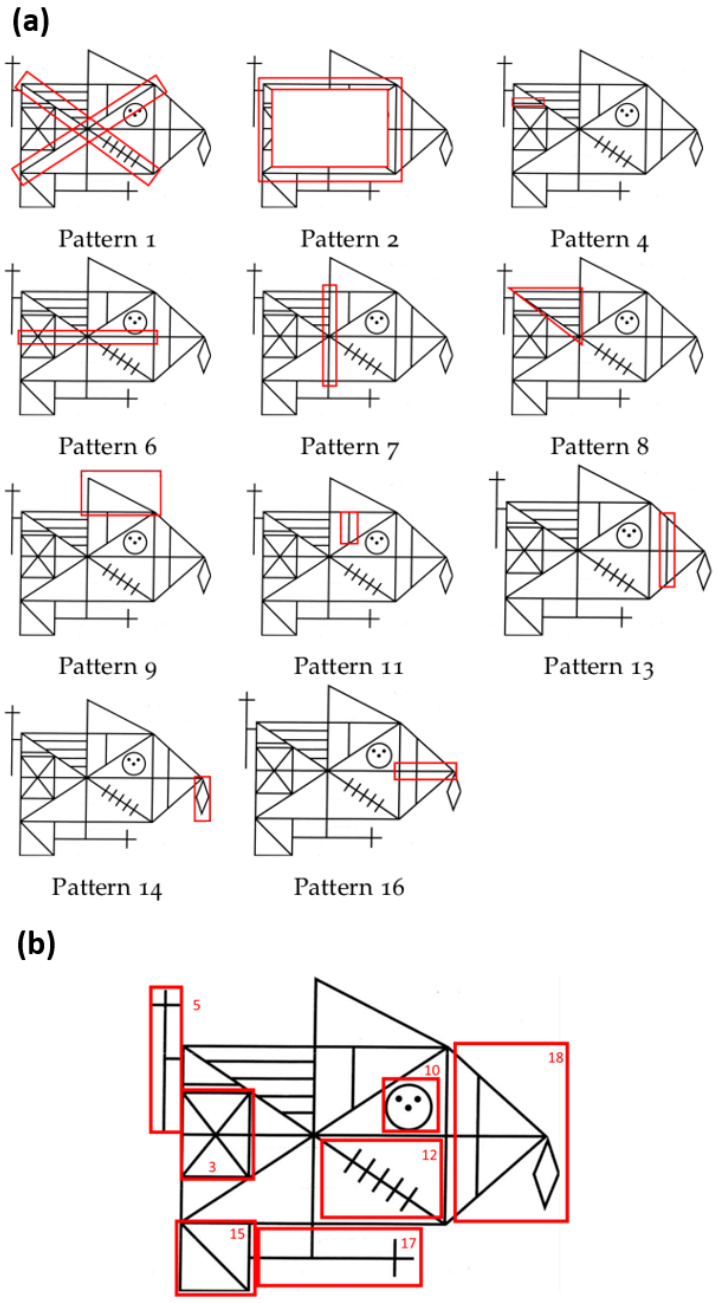

Figure 6.5: The ROCF patterns. Panel (a) shows the simple patterns and panel (b) shown the complex patterns. The red squares indicate the region of interest of each pattern.

were considered (the parameters $m_{\overrightarrow{clk}}$ and $m_{\overleftarrow{clk}}$ determined the clockwise or counterclockwise inclination respectively). The pattern were iteratively searched by moving the ROIs in the horizontal and in the vertical direction. In every iteration, A ROI was shifted by m_{sh} pixels, until it reached the image limits or additional pattern-specific bounds.

In the diagnosis formulation stage, the labels assigned to each pattern in the previous stage were used as features for the classification of the ROCF copy in one of the 3 clinical outcomes, thus determining the subject's most probable diagnosis.

6.2.5 Simple Patterns Evaluation

The simple patterns were characterised by an easy and regular geometrical structure, as a line, a regular polygon or multiple similar elements. CV algorithm were used for the automatic detection and evaluation of these patterns. The presence or absence of a simple pattern was determined using *line detection* algorithms, for those as patterns number 1, 4, 6, 7, 8, 11, 13 and 16, which were mainly composed by lines, and *shape detection* techniques, for the numbers 2, 9 and 14, which consisted in shapes. When a pattern was detected within a ROI, it was evaluated with *topological analysis*.

Line Detection

Line detection was performed using the probabilistic Hough transform algorithm implementation available in the OpenCV Python library [204]. The algorithm's parameters radius resolution r and the angle θ were fixed to 1 pixel and to $180/\pi$ respectively. The other parameters, namely the threshold for the minimum number of edges to detect a tract (th_e), the minimum line segment length (l_{min}) and the maximum number of gap between points allowed in a line (max_g), were set differently for each pattern. The algorithm returned the coordinates of all the detected lines edges. The lines belonging to the adjacent patterns and the other noise components were filtered out by setting a specific threshold (t_α) on their inclination w.r.t. the horizontal direction. Then, a set among the remaining lines was selected by counting the percentage of the pixels coinciding with those of the original pattern overlapping the sample ROCF and the template model. How much a collection of detected segments had to cover the template pattern to be identified as a pattern was decided by defining a threshold on such percentage (A_c). From the chosen set of segments, a single line was approximated using linear regression with the edges of all the lines in the set. In the pattern 8, composed by more than one line, the procedure was repeated for each line separately.

Shape Detection

For shapes recognition, the contour detection method in OpenCV [205] was used to find a set of geometrical objects in a ROI. To approximate the obtained contours to a regular figures, the OpenCV implementation of the Douglas-Peucker's iterative end-point fit algorithm [206] was used. Given a curve composed of line segments, a similar closed curve with fewer points was returned. Then a single shape was selected applying the following criteria: i) the convex shapes with the same number of vertices of the pattern to search was checked, evaluating the convex hull [207]; ii) if more than one shape met the requirements of (i), a threshold P_{min} was set to discard all the shapes whose perimeter was inferior to such value; iii) if there was still more than one eligible shape, a hierarchical representation of contours was computed and the innermost node (i.e. the outermost shape) was retained. The contours hierarchy was calculated using the OpenCV dedicated tool.

Topological Analysis

The label was assigned to a pattern in the sample ROCF according to the following rules: i) if no line or shape was detected in a ROI, the score *omitted* was assigned to that pattern; ii) if some shape was found, but none with the right number of vertices, the score *distorted* was assigned to the convex shape with a perimeter longer than P_{min}; iii) otherwise, the detected object, defined by the set of its edges coordinates, was topologically analysed using the the Python library Shapely [19]. The object was inspected to check if it satisfied a set of structural properties (as continuity, intersections and gaps) to be considered a *correct* drawn pattern. If the properties were partially satisfied the pattern could be labelled whether *misplaced* or *distorted*. Buffers were considered to account for the natural inaccuracy of hand-made drawing. Therefore, the thickness of the lines and shapes was increased according to two parameters, d_v for each vertex and d_l for each segment, as 'margin error', in pixels. The pattern-specific properties to be satisfied in the topological analysis are listed here, in the same order of pattern search.

- Pattern 2 *(main rectangle)*. The main rectangle was identified by 4 of the reference points selected by the user. The topological analysis returned: *correct*, if rectangular shape was found; *distorted*, if no rectangular shape was found.
- Pattern 1 *(main diagonals)*. $\mathrm{ROI}^1{}_0$ was split in four part, each of them was dedicated to the detection of one half of a diagonal. The topological analysis returned: *distorted*, if at least one half per diagonal has been detected or at least two halves of the same diagonal did not intersect; *misplaced*, if none of the previous conditions applied and at least one vertex did not intersect the corresponding vertex of pattern 2; *correct*, if none of the previous conditions applied.

- Pattern 4 *(short horizontal line)*. The topological analysis returned: *distorted*, if more than one lines was detected; *correct*, if a single line was detected.

- Pattern 6 *(horizontal line)*. If pattern 2 was previously found, the ROI^6_0 was centred with its horizontal axis of symmetry. The topological analysis returned: *distorted*, if a line was detected and at least one of its vertices did not intersect an edge of pattern 2; *misplaced*, if none of the previous conditions applied and the line did not intersect pattern 1 in a single point; *correct*, if none of the previous conditions applied.

- Pattern 7 *(vertical line)*. If pattern 2 was previously found, the ROI^7_0 was centred with vertical axis of symmetry. The topological analysis returned: *distorted*, if a line was detected and at least one of its vertices did not intersect an edge of pattern 2; *misplaced*, if none of the previous conditions applied and the line did not intersect pattern 1 in a single point; *correct*, if none of the previous conditions applied.

- Pattern 8 *(parallel lines)*. Pattern 8 was composed by four parallel segments so the initial ROI was divided in four sub-parts. The procedure was repeated iteratively by searching a segment in each sub-ROI. When a line was detected, the respective sub-ROI was excluded in the next iteration. The topological analysis returned: *distorted*, if less or more than four lines were detected and they did not intersect each other; *correct*, if none of the previous conditions applied.

- Pattern 9 *(topmost triangle)*. Pattern 9 was composed by four parallel segments so ROI^9_0 was divided in four sub-parts. A segment was iteratively searched in each sub-ROI. When a line was detected, the respective sub-ROI was excluded in the next iteration. The topological analysis returned: *distorted*, if a convex shape with more than 3 edges was found; *misplaced*, if none of the previous conditions applied and the shape had no intersections with other lines of the figure; *correct*, if none of the previous conditions applied.

- Pattern 11 *(short vertical line)*. The topological analysis returned: *distorted*, if more than one line was detected or one line was detected but it did not intersect either patterns 1 and 2; *misplaced*, one line which intersect either patterns 1 and 2 is detected and it intersect the two diagonals in the same point; *correct*, if none of the previous conditions applied.

- Pattern 16 *(right horizontal line)*. The topological analysis returned: *distorted*, if more than one line was detected or one line was detected but it did not intersect patterns 2 and 6; *misplaced*, if none of the

Table 6.2: Parameters setting of the computer vision algorithms for the detection and evaluation of the simple patterns.

	Definition of the ROI variants			Line/shape detection						Topological analysis	
Pattern n°	m_{sh} *[pixel]*	$m_{\overrightarrow{dk}}$ *[◦]*	$m_{\overleftarrow{dk}}$ *[◦]*	th_e	l_{min} *[pixel]*	max_g	t_α [◦]	A_{cov} *[%]*	P_{min} *[pixel]*	d_v *[pixel]*	d_l *[pixel]*
1	20	30	30	50	40	20	20-60	60	-	15	1.5
2	-	-	-	50	40	20	-	-	500	30	-
4	10	5	5	40	20	5	<10	60	-	20	1.5
6	15	10	10	75	40	20	<10	80	-	15	1.5
7	15	15	25	75	40	20	>70	80	-	15	1.5
8	5	5	12	30	10	5	<15	30	-	15	1.5
9	-	-	-	-	-	-	-	-	200	25	3
11	10	10	30	50	20	10	>60	60	-	15	1.5
13	10	10	30	50	20	10	>80	50	-	15	2
14	-	-	-	-	-	-	-	-	100	20	1.5
16	10	10	30	45	60	50	<10	80	-	15	3

previous conditions applied and the line did not intersect the manual rightmost point; *correct*, if none of the previous conditions applied.

- Pattern 13 *(right vertical line)*. The topological analysis returned: *distorted*, if more than one line was detected or one line was detected but it did not intersect at least one inclined edges of the delimiting triangle; *misplaced*, if none of the previous conditions applied and the line intersected the pattern 16 in its rightmost half; *correct*, if none of the previous conditions applied.

- Pattern 14 *(rhombus)*. The topological analysis returned: *distorted*, if a convex shape with more or less than 4 vertex was detected; *misplaced*, if none of the previous conditions applied and the shape did not intersect the rightmost manual point. *correct*, if none of the previous conditions applied.

The particular parameter choice for the CV algorithms used for the detection of simple patters is reported in Tab. 6.2, for each pattern.

6.2.6 Complex Patterns Evaluation

The complex patterns were detected using DL algorithms by finding the most similar region in the ROCF sample image to the pattern in the template model. Then they were evaluated using a measure of similarity and topological analysis.

The concept of similarity was defined as the euclidean distance (L_2) between two images, calculated by mapping them into a 1024-dimensional embedding space[208]. Two images were more similar the shorter the L_2 distance between their vectors in the embedding space. An embedded representation of the images was obtained using a modified ResNet50V2 neural network architecture [29]. The fully-connected layer on top of the network was removed, the output of the residual part flattened and a 1024-rectified linear units (ReLU) fully-connected layer was added. A triplet loss as cost

function was chosen, which has been shown to be efficient for this type of tasks [208]. The triplet loss, Eq. 6.1, encouraged the images of the same pattern to be projected onto very close points in the embedding space and it enforced the margin between images of different objects by considering triplets of vectors.

$$L = max(m + D(\xi_a, \xi_p) - D(\xi_a, \xi_n), 0). \tag{6.1}$$

In the equation of the loss, ξ_a was the vector of a reference image (the *anchor*) in the embedding space, ξ_p was the vector of an image of the same object of the anchor (the *positive*) and ξ_n was the vector of an image of a different object (the *negative*); $D(\xi_i, \xi_j)$ was the squared euclidean distance between the vectors of i and j; and m was the margin between positive and negative pairs. The loss minimisation must satisfy the constraint in Eq. 6.2,

$$D(\xi_a, \xi_p) + m < D(\xi_a, \xi_n), \tag{6.2}$$

therefore, $D(\xi_a, \xi_p)$ was pushed to zero and $D(\xi_a, \xi_n)$ to be grater than the former plus m.

A different network was trained for the recognition of each complex pattern. Pattern-specific dataset were created by manually cropping the pattern representations from the ROCF of all the subjects. For the similarity measurement, *correct* and *distorted* patterns only were considered. The scarce amount of samples retrieved was incremented ten times applying the following data augmentation techniques to each dataset. The Phyton library ImgAug [209] was used to perform the following image transformations in random order:

- Gaussian blur with variance ranging from 0 to 0.5,
- aspect ratio preserving scaling with a factor ranging from 0.85 to 1.15,
- rotation by -10 to 10 degrees,
- shear mapping by -15 to 15 degrees,
- translation by -40 to 40% on x-axis and y-axis independently.

However, datasets were strongly unbalanced as the portion of wrongly drawn patterns was consistently lower. In particular the percentage of correctly executed patterns was 66%, 77.4%, 86.7%, 70.8%, 88.8%, 72.5% and 70.2% for patterns 3, 5, 10, 12, 15, 17 and 18 respectively. Therefore, heavier forms of image augmentation to 100 correct samples to obtain other 1000 'distorted' versions were performed, with the aim of re-balancing the datasets. The new data augmentation parameters were manually set to better resemble the 'true' distorted samples. To train a network, a batch of 32 pattern images was randomly extracted from a dataset and each of them was paired

with the same pattern of the template model. A positive pair was generated if the image was labelled as *correct*, or a negative pair if it was labelled *distorted*. The network computed the L_2 distance between the images of each pair and the triplet loss was computed using the batch-hard strategy [210], i.e. by selecting the hardest positive pair (with the maximum L_2) and the hardest negative pair (with the minimum L_2) only. The training consisted in 50 epochs in which the dataset was split in training and validation/test set by randomly picking the 30% of the samples from both the correct and distorted class. The Adam optimiser [211] was used with a learning rate of 0.0001 and the model with the lowest validation loss value among each epoch was retained.

For detecting a complex pattern, the set its initial ROIs was moved in the vertical and horizontal direction for a maximum distance of 50 pixels. For each pattern, the ROI with the maximum value of similarity w.r.t to the template was retained. Then, the assignment of the label according to the followed procedure:

- *omitted*. A linear support vector machine (SVM) classifier was trained and used to discriminate between the presence or the absence of the pattern counting the portion of non-white pixels contained in the ROI. Pattern-specific dataset were used, including the cropped samples of all the omitted and the non-omitted patterns. The balanced accuracy of the classification was estimated with the leave-one-out cross-validation (LooCV); The label *omitted* was assigned when the SVM returned the absence for the pattern.

- *misplaced*. If a pattern was detected, topological analysis was applied to examine its correct location by checking all expected intersections with the surrounding patters;

- *distorted* or *correct*. If the pattern was neither omitted nor misplaced, a second linear SVM classifier was trained and used to discriminate between a distorted and a correct representation. Pattern-specific dataset, with correct and distorted patterns only, were used as training data and the L_2 from the template pattern was considered as single input feature. The balanced accuracy of the classification was estimated with LooCV.

6.2.7 Diagnosis Formulation

The diagnosis formulation stage was aimed at associating the most probable subject's diagnosis (between *healthy*, *MCI* and *dementia*) to each ROCF sample, using the 18 labels assigned to its patterns as predictors. The ability

of the system to discriminate between groups was investigated by setting various classification tasks. Four classifications were considered: *healthy-MCI*, *healthy-dementia* and *MCI-Dementia*; and a multi-class classification including the 3 groups. Since the classes were unbalanced, the four classification tasks were repeated with new datasets created by randomly sampling (without replacement) 50 elements per class and by averaging the outcomes of 50 iterations, for a more robust estimate. A state-of-art boosting algorithm, Catboost [212], was trained to solve the classification tasks by choosing a weighted cross entropy loss function [213] and setting a number of 500 iterations. The performance were evaluated by estimating the Accuracy, F1, Precision and Recall scores with the Leave-one-out cross-validation (LooCV). The normalised numerical labels of the patterns (0, 1, 2, 3) were normalised (between 0 and 1) and used as input features. The classification tasks were implemented with the Python library SciKit-learn [176].

6.2.8 Model Explanation

The binary classification tasks were further analysed by applying the model explanation technique SHAP [214, 214]. Here, 3 randomly sampled datasets with 50 samples per class were used, to avoid the stronger influence of the more numerous group in the classification. SHAP used game theory to rank the features (i.e. the patterns) importance and to assess the contribution of each feature in the binary classification of a ROCF sample in the trained models. The single feature contribution in the classification of each sample was quantified by a weight (the Shapely value) which moved its prediction toward a class or the other, if negative or positive, by an amount proportional to its magnitude. The features rank was then obtained by considering the average absolute weight of each feature for each sample. With this technique, one could better interpret the effect of each single pattern in the binary classification tasks and thus appreciate the sensitivity of the patterns in discriminating between healthy from pathological individuals and between different levels of cognitive decline.

6.3 Results of the DSS for the ROCF Analysis

6.3.1 Simple and Complex Patterns Evaluation

For each simple pattern, the total evaluation Accuracy was calculated as the percentage of the correctly labelled patterns over the total number of patterns. The total accuracy scores for the simple pattern are reported in the upper section of Tab. 6.3. For the evaluation of the complex patterns instead, a first SVM model predicted the presence or absence (label *omitted*) of a pattern in the ROCF sample. Than, in the case of presence, topological analysis was used to determine whether the pattern was *misplaced* or not. If

Table 6.3: Pattern scoring accuracy.

Simple patterns	total acc. [%]		
1	65.5		
2	66.8		
4	72.2		
6	78.9		
7	77.6		
8	68.2		
9	61.9		
11	72.2		
13	71.3		
14	68.2		
16	74.9		
Complex patterns	omitted/others [balanced acc.]	correct/distorted [balanced acc.]	total acc. [%]
3	0.94	0.81	61.4
5	0.81	0.76	61.9
10	0.91	0.77	70.0
12	0.83	0.86	68.2
15	0.89	0.89	65.9
17	0.71	0.82	77.1
18	0.65	0.82	68.2

the pattern was present and not misplaced, a second SMV model predicted whether the pattern was *distorted* or *correct*. The lower section of Tab. 6.3 shows the balanced Accuracy of the SVM models and the total Accuracy for each of complex patterns. The total Accuracy score corresponded to a 4-element labelling task and lower accuracy scores were observed for the patterns 3, 5 and 9 (61.9%, 61.4% and 61.9% respectively). Patterns 6, 7, 16 and 17 achieved the highest total Accuracy scores (78.9%, 77.6%, 74.9% and 77.1% respectively) and the rest of the patterns gained and average total Accuracy score of 68.8%.

6.3.2 Diagnosis Classification

The 4 classification tasks (healthy-MCI, healthy-dementia, MCI-dementia and 3-class) were performed using all the samples contained in each class. Then, the same tasks were executed using new dataset created by randomly sampling 50 samples for each class. In the latter case, the performance metrics were averaged from 50 iterations of dataset creation and classification. The Accuracy, F1, Precision and Recall scores of all the tasks are reported in

Table 6.4: Results of the diagnosis classification tasks.

	All samples				Random sampling			
	healthy-MCI	healthy-dementia	MCI-dementia	3-class	healthy-MCI	healthy-dementia	MCI-dementia	3-class
Accuracy	0.89	0.92	0.87	0.76	0.85±0.03	0.91±0.02	0.83±0.03	0.73±0.04
F1	0.93	0.91	0.79	0.74	0.85±0.04	0.90±0.03	0.84±0.03	0.73±0.03
Precision	0.95	1.00	0.85	0.74	0.89±0.04	0.96±0.02	0.84±0.04	0.74±0.04
Recall	0.90	0.84	0.74	0.74	0.81±0.05	0.85±0.04	0.89±0.04	0.73±0.04

Tab. 6.4. In the left panel are listed the outcomes of the classifications made using all the samples of each group. In the right panel, the average outcomes (with standard deviation) are shown for the tasks performed with random sampled datasets. High to excellent performance were obtained in the binary classification tasks with all the samples, with Accuracy scores ranging from 87% (MCI-dementia) to 92% (healthy-dementia); F1 between 79% (MCI-dementia) and 93% (healthy-MCI); Precision between 85% (MCI-dementia) and 100% (healthy-dementia); and Recall between 74% (MCI-dementia) and 90% (healthy-MCI). Similar performance were achieved mediating the outcomes of the classifications with the random sampled dataset. The lower performance were obtained in the more complex multi-classification tasks: scores from 73% to 76% were obtained for the Accuracy and scores from 73% to 74% for the other metrics.

The confusion matrices and the receiving-operator curves (ROC) for the 4 classifications in which all the samples were retained are shown in Fig. 6.6. Results showed the high ability of the system in discriminating between healthy and pathological individuals or between a mild and a more severe cognitive decline, when binary classification problems are considered. Therefore, the DSS was capable to retrieve a significant amount of information through the automated patterns analysis. The 3-groups classification achieved lower scores of performance. However, the limitations are linked to the increased complexity of the task and the limited amount of data available.

6.3.3 Model Explanation

The model explanation technique SHAP was applied to 3 binary classification tasks, i.e. healthy-MCI, healthy-Dementia and MCI-dementia, considering 3 randomly sampled dataset with 50 samples per class. The outcomes revealed how the features (i.e. the ROCF patterns) ranked by importance and how their value (i.e. label) contributed in the classification of the samples in the tasks. The average absolute Shapely values of each pattern are displayed in Fig. 6.7, column (a), for the 3 tasks. The SHAP analysis revealed that patterns 9, 11, 3, 6 and 4 were the most sensitive in the discrimination between healthy subjects and MCI patients (as their average absolute Shapely value was greater). Alternately, in the classification between MCI and dementia patients the most informative resulted the patterns number 10, 13, 3, 18 and

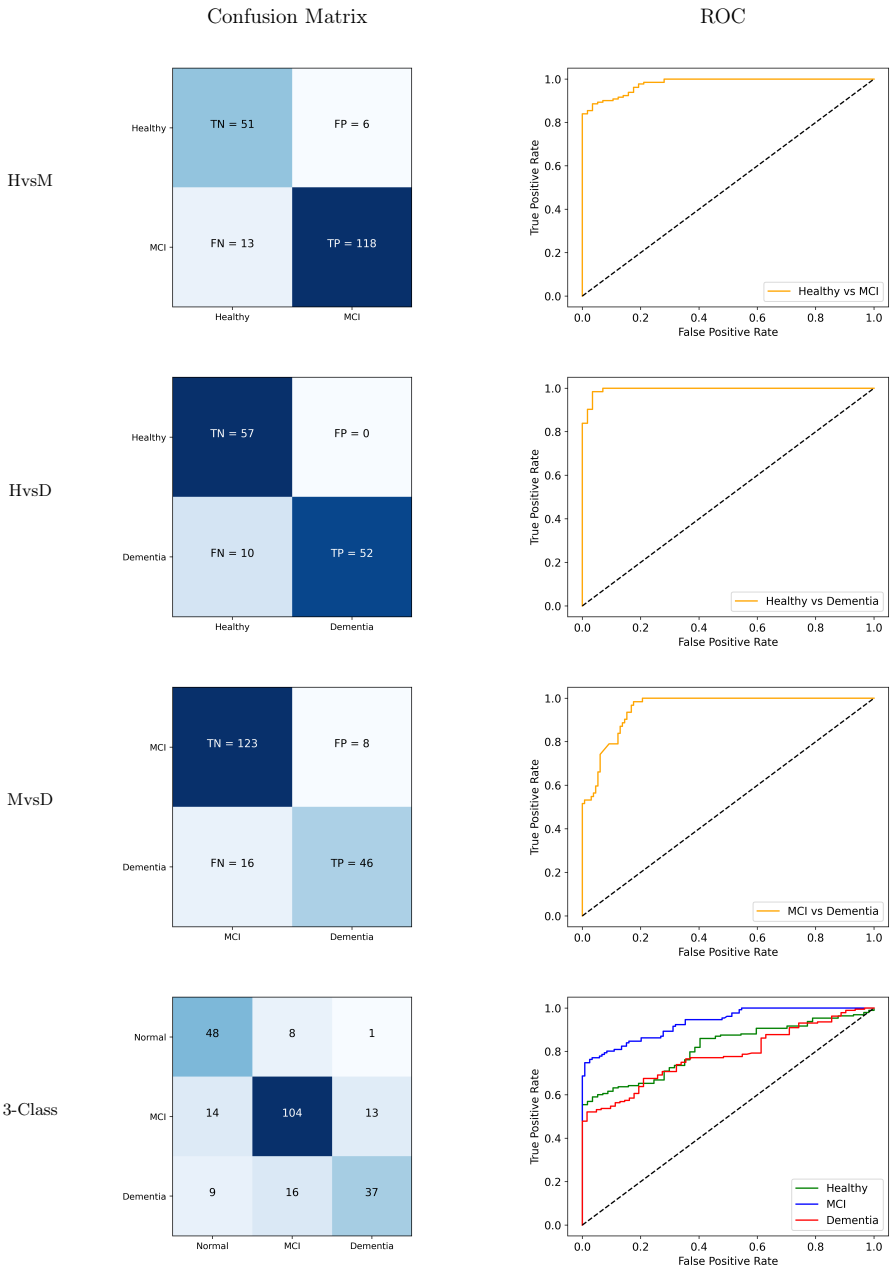

Figure 6.6: Performances for the 3 binary and the multi-class classification tasks performed with all the samples (healthy vs MCI in the first row, healthy vs dementia in the second, MCI vs dementia in the third). The confusion matrices are reported in the left column and the ROC curves in the right column. For the 3-class classification, the ROC curves are figured in the same plot in different colours: green for the healthy class, blue for the MCI class and red for the dementia class.

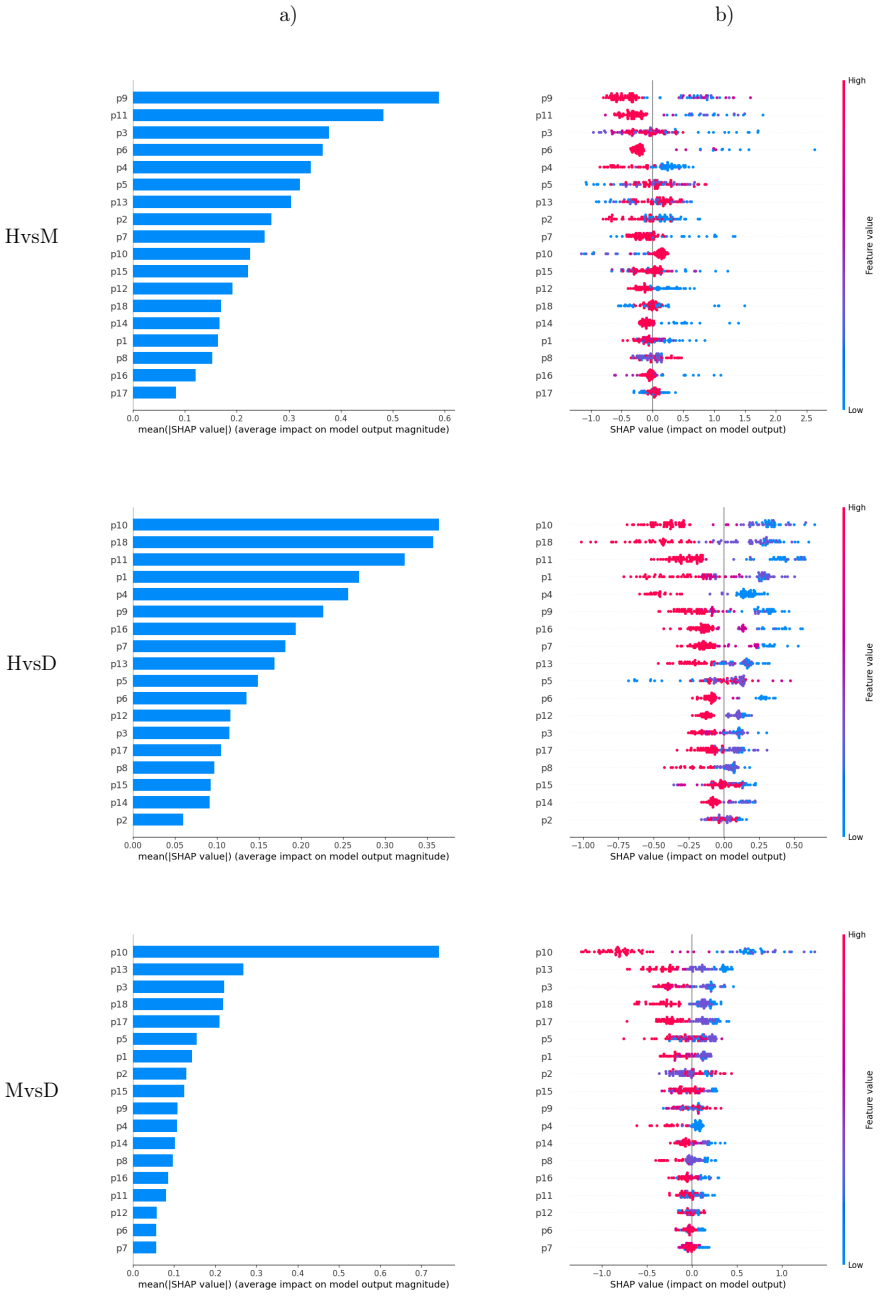

Figure 6.7: Shapely values for the binary classifications of the ROCF samples (healthy vs MCI in the first row, healthy vs dementia in the second, MCI vs dementia in the third). Average Shapely values in column (a), scatter plots of shapely values in column(b).

17. In the remaining task (healthy-Dementia), similar patterns to both the former cases were found as the most sensitive. The plots in Fig 6.7, column (b), show the single Shapely values of the patterns for the classification of each sample. Negative Shapely values pushed the classification of a sample towards the 'less severe diagnosis' (i.e healthy in the first two tasks and MCI in the third), while positive values moved the prediction towards the other diagnosis. The blue-red colour-map of the dots indicated the quality of the pattern representation, as a continuous interpolation from the minimum score of 0 (in blue), corresponding to the label *omitted*, to the maximum score of 3 (in red), which corresponded to the label *correct*. The use of this model explanation technique was useful to reveal how the sensibility of the pattern can change according to the level of the cognitive decline in the individual. Moreover, it gives information about the type of error a group of individuals tend to commit more in the representation of the patterns.

6.4 Discussion

In this study, a DSS for the analysis of the ROCF copy test has been proposed to assist clinicians in the diagnosis of MCI and dementia. The system performs a semi-automatic analysis of a scanned ROCF image, returning a qualitative score for each pattern and the most probable clinical outcome for the subject between healthy, MCI or dementia.

To perform the analysis, the clinician is asked to select 5 reference points in the figure. They indicate the main rectangle edges and the rightmost vertex of the triangle on the right, which may often result very confused in some distorted ROCF copies. Therefore a decision by the expert is required. The DSS evaluates the patterns using both computer vision (CV) and deep learning (DL), calibrated and trained using retrospective data of past ROCF tests from healthy, MCI and dementia individuals. Then, it formulates the most probable diagnosis using classification methods. The ability of the system in discriminating between different diagnosis was assessed by testing different classification tasks.

In addition, explainable AI was used to interpret the models decision by estimating the single contribution of the patterns in the determination of a subject's diagnosis. This last DSS feature offers further insights on the patterns sensibility to different levels of severity of mental decline.

The pattern evaluation phase of the DSS aimed at replicating an expert-based qualitative labelling. The patterns of each sample in the dataset were manually labelled as *omitted*, *distorted*, *misplaced* and *correct*. The four labels were coded with integer numbers reflecting the quality of the pattern representation, from 0 to 3.

The simple patterns were evaluated using hard-coded rules and low to

medium total Accuracy scores were obtained in their labelling. The main reason behind the choice of CV methods for the evaluation of the simple patterns consisted in the small amount of samples available for each label category. In some cases indeed, the rules were adjusted on a limited examples of differently labelled patters. For example, pattern 9 resulted the hardest to be evaluated (with a total Accuracy of 61.9%), since only a small portion was labelled as *distorted* and a very diverse group of representations fell into that category. The variety of the patterns replicas represented the main issue to the formulation of general rules for the exact labelling (Tab. 6.1), although some positive total Accuracy scores were obtained in the evaluation of the patterns 6, 7 and 16 (Tab. 6.3).

For the evaluation of the complex patterns, the articulated shapes represented a further barrier to the use of explicit labelling rules. Therefore, machine learning models were applied in their evaluation. A first SVM model recognised the presence of a pattern with high Accuracy scores, ranging from 70% to 94%. Pattern 18 was the hardest to be detected by the SVM, with an accuracy of 65%. Then, DL models were trained to estimate a measure of a pattern 'good representation' based on their similarity w.r.t. the template model. The decision between the *correct* and the *distorted* label was determined by a second SVM model, which achieved good Accuracy scores (ranging from 76% to 89%). Only the assignment of the label *misplaced* was implemented using explicit rules. The total Accuracy scores were similar to those achieved with the simple patterns. Since the number of the differently labelled patterns was unbalanced, data augmentation was used. However, it could have resulted a too strong approximation of the real patterns representation by the subjects. Yet it is worthy to highlight that the total Accuracy scores of the pattern evaluation were computed as the ratio of those correctly and wrongly labelled, considering the 4 four label categories. It is also known that the combination of weak classifiers can handle noisy datasets and outperform strong learners [215].

Very good to excellent classification metrics were obtained in the discrimination between all the healthy and MCI individuals (Tab. 6.4). Despite of the class unbalance, high Precision and Recall were achieved (95% and 90% respectively). The ROC analysis revealed that the threshold of the classifier could be adjusted to calibrate the true positive rate to the maximum of 100%, tolerating a false positive rate of 30%. Even higher performances were achieved in the classification between healthy and dementia individuals, with the 100% of Precision score and a Recall equal to 84%. Such behaviour was quite expected since an increased level of impairment of cognitive functions is present in patients with dementia. Slightly lower performance were obtained in the MCI-dementia classification with all the samples. However, the initial Recall of 74% could be improved to 100% by selecting a different threshold with a Precision equal to 80%. The fourth classification, with the 3 groups, presented the lower performances (still good with 76% for the

accuracy and 74% for the other metrics) as the task complexity increased. The same tasks were repeated using random sampled balanced datasets and mediating the classification metrics of 50 iterations, for a more robust performance estimation. A higher Precision was kept in the tasks healthy-MCI and healthy-dementia, while Recall resulted higher in the MCI-dementia classification. Similar performance were kept in the 3-class task.

In the binary tasks the contribution of the patterns in discriminating between different diagnosis was further investigated using a model explanation technique. This analysis was aimed at identifying which patterns resulted as the more sensitive in the recognition of the different severity levels of cognitive decline (i.e MCI and dementia) and at revealing how much the quality of the pattern representation weighted in the classification of each specific class. Result showed different sets of patterns among the most sensitive in the discrimination between the healthy-MCI and the healthy-dementia groups. For example, patterns 9, 11, 6 and 4 were among the most important in the former task, while they resulted in lower ranking positions in the latter. In the task MCI-dementia instead, the most informative patterns were the numbers 10, 13, 3, 18 and 17. In the classification between healthy and dementia, some of the most significant features corresponded to those of the other two tasks. As clearly visible in the scatter plots (Fig. 6.7, row (b)), the Shapely values of the most important features had a sharp separation between the high and the low features values in the two classes and the dots were quite spread in the horizontal direction. A higher feature value indeed represented a better quality (a better label achieved) of the pattern representation. Consequently, higher feature values resulted more distributed on the left side of the plots, i.e. in the class of those individuals with a less severe cognitive decline. However, the separation between a good and a bad pattern representation could appear not so clear in some cases. For example in the task healthy-MCI, there was not an evident distinction between the two classes in the labels assigned to patterns 5, 13, 15, 18 , 8 and 17. It could suggested both that healthy subjects may draw those pattern incorrectly or misplaced, and that MCI patients may be able to correctly represent them. In the task MCI-dementia, the higher and lower feature values were not well separated in the case of patterns 2, 9, 11 and 12. This may imply that misplacement or bad representations of those patterns were likely to be seen in both groups. In the task healthy-dementia, in which the gap in the cognitive functionality is more marked, a larger separation between high and low feature values appeared for almost all the patterns.

6.5 Clinical Impact of the Expert System

The DSS presented in this work is a prototype which can be used in the study and development of medical expert systems applied in the diagnosis of cognitive decline. It was a retrospective study in which past data samples of the ROCF copy test were collected and accurately inspected by trained operators to build the system's base knowledge. The amount of past ROCF examples constitutes the information level of the expert system, therefore it is likely to increase as the number of data samples grows. An higher level of information enhance the accuracy of the DSS in the pattern evaluation and the confidence in the formulated diagnosis. Currently, samples data consist of domain experts supervised ROCF copies, combined with the patient's true diagnosis. However, larger amount of samples might also allow the exploitation of unsupervised techniques for the image analysis. Thus removing any residual subjective component in the evaluation of the test.

Although explainable AI methods are applied to better interpret the DSS proposed diagnosis, the system is not meant to take critical decisions in the clinical setting. Indeed, the tool is proposed as a decision support system to assist clinicians in the evaluation of the ROCF copy test by supplying additive and objective information.

The extensive use of the system in the clinical practise might improve the inter-rater reliability in the evaluation of the test. With this objective, longitudinal control studies could be addressed to assess whether the use of the DSS in clinical practice would enhance accuracy and efficiency in the process of the cognitive evaluation. In particular, it could help to more objectively define a neuropsychological sign at the test, contributing to a more reliable assessment. In addition, the DSS approach to neuropsychological testing may help to discover and define new signs of cognitive dysfunction, thus directly assuming high clinical value.

By relying on sensitive measures, it could help detect a neuropsychological sign earlier than when it is typically done in the clinic, helping to place a diagnosis at an earlier stage in the disease course. Indeed, the DSS allows the remote (and posterior) analysis of the ROCF copy test. In remote monitoring scenario, the expert system could be used by the practitioners to evaluate the ROCF drawings done by subjects during tele-consultations or in self-administered tests. Poor test results may suggest a possible degradation of cognitive function. This would represents a more agile solution to detect abnormal changes in the cognitive functionality of the individuals than in-persons visits.

Chapter 7

Conclusions

This PhD research proposed novel eHealth methodologies to support the early identification of age-related decline. Early diagnosis of chronic diseases in older adults is a key element for the adaptation of the current health-care models worldwide to face the rapid ageing population. Indeed, only well timed interventions can stop or significantly slow down the effects of a pathological decline in seniors, which leads to a poor quality of life for the patient and a large impact on the public care system. Remote monitoring of community dwelling elders and the remote assessment of the physical/cognitive functionality are possible strategies to improve the probability to early diagnose a chronic disease. When in a pre-clinical stage, the typical symptoms of the chronic diseases may not yet manifest. Therefore, a continuous tracking of the subject's health status could allow the detection of some anomalous variation in his physical/cognitive functionality which may suggest an abnormal ageing process. Furthermore, tele-consultations and remote evaluation of some clinical tests can provide operators with more precise information on the targeted health aspects of the subject, in a more agile and efficient fashion than the in person visit. This work proposed a solution for the remote monitoring of the older adults physical and cognitive decline through the ecological assessment of handwriting and an expert system to help the experts in the remote examination of the Rey-Osterrieth complex figure (ROCF) test, which is an important tool for the diagnosis of MCI and dementia.

Ecological Assessment of Handwriting

The first objective consisted in a transnational study in which the application of the handwriting analysis, as biomarker for the age-related decline, was investigated and adapted in the unprecedented context of remote monitoring. Being a common exercise and having been observed to vary with ageing and pathology, handwriting appeared as an optimal candidate for the instrumental daily-life activity monitoring purpose. The research was

characterised by the following approach: first, i) a deep investigation of the handwriting assessment in literature was carried with the aim of finding the current methodologies, experimental settings, results and limitations for a remote monitoring application; Then, ii) the limitations were addressed and overcome with the design of a novel smart object technology and the application cutting-edge methods from data analysis and explainable artificial intelligence; iii) each design and development stage of the proposed solutions were tested using ad-hoc experimental protocols and real-world data.

As explained in Chapter 3, the handwriting assessment consisted in the analysis of certain indicators, computed from the signal acquired during the handwriting activity. These quantities, related to the temporal, dynamic and tremor characteristics of the writing gesture, were studied to undergo variations with the presence of age or pathological decline in the individuals. However, some limitations were found for the application of the handwriting assessment in the remote monitoring in the home environment. A first barrier was represented by the data acquisition instruments (tablets or digitizers), which were not easy to be independently operated by older users. Then, the use of standard protocols for the data acquisition in literature did not allow to extend the results of the previous studies in the uncontrolled domestic environment. Indeed, the use of standard protocols was excluded in the remote monitoring scenario for the lack of any supervision.

The solutions to overcome the limitations related to the data acquisition devices and the ecological validity of the handwriting assessment were explained in Chapter 4. A sensorized IoT ink pen was developed to allow the ecological handwriting gesture data acquisition in paper-and-pen tasks. The device was successfully tested, validated and then used to collect data from healthy subjects. Then, data were used to study the handwriting assessment in a conditions which was similar to the uncontrolled remote setting. Indeed, the data acquisition protocol was designed to resemble the common writing activity: subjects were asked to write a short free text and a grocery list, without any further constraint. As first, the reliability of the indicators was confirmed in test-retest writing tasks. Then, their sensibility to age-related variation was studied in 3 differently aged groups of healthy subjects. In a total of 12 indicators, 8 of them showed significant changes with age. This result confirmed the possibility to use the handwriting assessment as an instrument to detect variation related to the ageing process in uncontrolled environments. Although the variation trend of each indicator with age was analogous with those reported in literature, their mean values not always corresponded with the previous studie in controlled settings. The reason should be found in the experimental setting used in this work, which was designed to be minimally constrained and the most similar to the normal writing activity. This rather uncontrolled framework represented a substantially different setup than the previous ones and the measurements were more likely to be affected by noise. As a consequence, a remote monitoring system

based on the handwriting assessment should not rely on standard reference values for the indicators, but it should consider a multivariate approach in which various patterns of indicators may be associated to age or pathological variations in the subject.

In Chapter 5, a supervised and population-specific anomaly detection approach was presented as a method to exploit the handwriting assessment in the remote monitoring of the subject's age or pathological-related decline. The aim was that to use the handwriting indicators to classify subjects in the correct age class or pathological (PD) group they belonged. In the remote monitoring context, the unexpected classification of an individual (for example a subject aged between 60-70 classified as an over 70 years of age, or an healthy subjects classified as a PD patient) may indicate an abnormal ageing process. Therefore, such event should trigger a deeper investigation on his health status. This strategy was tested by solving various classification tasks between group of differently aged healthy individuals and a group of PD patients. A state-of-art classification algorithm (Catboost) was used and the achieved performance were high in terms of Precision and Recall in all the tasks. Since the classification algorithm was a black box model, a recent explanation technique (SHAP) was applied to understand each model decision. This additional method increased the interpretation of the anomaly detection strategy, as it revealed the impact and the behaviour of the handwriting indicators (known quantities) in the identification of the subject's group.

The proposed solution combined the more traditional signal processing techniques with the novel advancement in artificial intelligence to maximise the information obtainable from noisy and multidimensional data. Raw handwriting data were analysed using conventional techniques to obtain reliable quantities with a precise *physical* meaning. This allowed the association of the handwriting measurements to specific behaviours in the gesture dynamic and tremor domains, and it also include the possibility to clinically interpret the indicators by healthcare professionals. Artificial intelligence was then used to efficiently search in the highly complex space of the handwriting indicators the patterns which may be related to the physical or cognitive decline in the subjects. The effectiveness of this approach is evidenced by the high performances obtained by the classification algorithms, which usually need a large amount of data to match such results. In addition, the use of explainable AI allowed to relate the outcomes of the classification algorithms to the *physically*-explicable domain of the handwriting indicators.

A critical point of this proposed solution can be found in the limited amount of data that could be collected (80 healthy subjects and 20 PD patients). Although the modest number of samples, very good outcomes in terms of classification performances were obtained. Yet, a more robust performance estimation would be envisaged using a larger number of individuals. Furthermore, the addition of more pathological categories (such as

MCI and dementia), would enlarge the spectrum of the abnormal detectable conditions. Another limitation can be addressed to the population-specific nature of the anomaly detection methods which was tested in this research. A more subject-centred approach would increase the sensibility in the identification of the abnormalities in the handwriting data, overcoming the problems related to the inter-subjects variability. However, longitudinal studies are necessary to acquire subject-specific data which may supply information about the individual's decline trajectory.

Expert System for the ROCF Copy Test

The second objective (Chapter 6) consisted in the design and development of a decision support system (DSS) for the analysis of the ROCF copy test. In the clinical cognitive assessment, the ROCF is one of the non-verbal tests which can be administered to the subject to investigate several cognitive functionalities. The specialists can use battery of tests to diagnose age-associated pathological conditions such as MCI and dementia, yet some issues related to the in-person visits (for example the examination is usually posterior to the evident manifestation of the symptoms in the individuals and the waiting times of outpatient facilities are generally long) paradigm may cause a delay in the diagnosis. To overcome this issue, various eHealth strategies have been investigated in literature to allow the remote administration or evaluation of some cognitive tests: they can be remotely supervised during tele-consultations or self-administered by the subject and forwarded to the specialist.

The application of the ROCF test in a remote monitoring setting could give important information about the cognitive decline in the subject, because of its clinical relevance. However, some studies pointed out that the interpretation of the test though the standard procedure may have a poor inter-rater reliability. Although it might not impact the diagnosis in the clinical practice (since a number of various cognitive tests are applied), this issue could affect the remote evaluation of the self-administered test by the subject and overlook the presence of fleeing signs related to a cognitive degeneration during time. To address this limitation an expert system based on retrospective knowledge have been proposed to support the clinicians in the evaluation of the ROCF copy test.

The DSS was based on retrospective knowledge collected from past ROCF copies examples from normal, MCI and dementia individuals. It used computer vision and deep learning algorithms to detect and evaluate the 18 patterns in the ROCF, by assigning a qualitative score each. Then, it used the scores to classify the image with the most probable diagnosis. The DSS was able to correctly discriminate the ROCF copy tests between healthy and MCI individuals, and between MCI and dementia patients, with good

levels of Precision. Domain specialist could use the system to have suggestions about the evaluation of the single patterns of the ROCF and about the most probable diagnosis for the examined individual, especially in a remote assessment framework. In addition, the information the specialists would retrieve by using the DSS was enriched with the application of the explainable AI model SHAP. This tool returned the impact of each of the 18 ROCF's patterns in the classification outcomes, so it revealed how the sensibility of the patterns changed with the progression of the cognitive decline in the individuals. The system could be also used in research, as it might highlight correlations of different mental disorders or impaired functionalities with particular elements in the ROCF copy test.

The DSS was based on retrospective knowledge, therefore it is likely to improve its accuracy and robustness as the number of past ROCF test examples increases. The limited number of samples in this study might have affected the accuracy in the evaluation of some patterns, yet the performance in the diagnosis classification were surprisingly high. Especially considering that the ROCF copy test is just a partial element of the clinical cognitive assessment procedure.

A very large collection of ROCF copies could improve the performance of the DSS in the pattern evaluation and increase the robustness in the classification of the pathology. Higher amount of data could also encourage the use or more unsupervised approaches in the analysis of the figure, which it may exclude any subjective component in the evaluation of the ROCF and increase its general validity.

Bibliography

[1] "World report on ageing and health," World Health Organisation, Tech. Rep. 978 92 4 069479 8, 2015.

[2] J. Rasmussen and H. Langerman, "Alzheimer's disease - why we need early diagnosis," *Degenerative Neurological and Neuromuscular Disease*, vol. Volume 9, pp. 123–130, 12 2019.

[3] S. Gordon, N. Baker, M. Kidd, A. Maeder, and K. Grimmer, "Pre-frailty factors in community-dwelling 40-75 year olds: Opportunities for successful ageing," *BMC Geriatrics*, vol. 20, 03 2020.

[4] L. Fried, C. Tangen, J. Walston, A. Newman, C. Hirsch, J. Gottdiener, T. Seeman, R. Tracy, W. Kop, G. Burke, and M. A. Mcburnie, "Frailty in older adults evidence for a phenotype," *The journals of gerontology. Series A, Biological sciences and medical sciences*, vol. 56, pp. M146–56, 04 2001.

[5] K. Langa and D. Levine, "The diagnosis and management of mild cognitive impairment," *JAMA*, vol. 312, pp. 2551–61, 12 2014.

[6] J. Valderas, B. Starfield, B. Sibbald, C. Salisbury, and M. Roland, "Defining comorbidity: Implications for understanding health and health services," *Annals of family medicine*, vol. 7, pp. 357–63, 07 2009.

[7] F. Lunardini, M. Luperto, M. Romeo, N. Basilico, K. Daniele, D. Azzolino, S. Damanti, C. Abbate, D. Mari, M. Cesari, A. Borghese, and S. Ferrante, "Supervised digital neuropsychological tests for cognitive decline in older adults: Usability and clinical validity study," *JMIR mhealth and uhealth*, vol. 8, 11 2020.

[8] J. Walton, "Handwriting changes due to aging and parkinson's syndrome," *Forensic science international*, vol. 88, pp. 197–214, 09 1997.

[9] B. Engel-Yeger, S. Hus, and S. Rosenblum, "Age effects on sensory-processing abilities and their impact on handwriting," *Canadian*

journal of occupational therapy. Revue canadienne d'ergothÃ©rapie, vol. 79, pp. 264–74, 12 2012.

[10] S. Rosenblum, P. Werner, T. Dekel, I. Gurevitz, and J. Heinik, "Handwriting process variables among elderly people with mild major depressive disorder: A preliminary study," *Aging clinical and experimental research*, vol. 22, pp. 141–7, 04 2010.

[11] M. Schmuckler, "What is ecological validity? a dimensional analysis," *Infancy*, vol. 2, 10 2001.

[12] F. Lunardini, M. Luperto, M. Romeo, J. Renoux, N. Basilico, A. Krpic, A. Borghese, and S. Ferrante, "The movecare project: Home-based monitoring of frailty," 05 2019, pp. 1–4.

[13] K. Zeuner, M. Peller, A. Knutzen, I. Holler, A. Munchau, M. Hallett, G. Deuschl, and H. Siebner, "How to assess motor impairment in writer's cramp," *Movement disorders : official journal of the Movement Disorder Society*, vol. 22, pp. 1102–9, 06 2007.

[14] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of parkinson's disease," *Artificial intelligence in medicine*, vol. 67, 01 2016.

[15] S. L. Hong, E. James, and K. Newell, "Coupling and irregularity in the aging motor system: Tremor and movement," *Neuroscience letters*, vol. 433, pp. 119–24, 04 2008.

[16] P. Spaan, J. Raaijmakers, and C. Jonker, "Alzheimer's disease versus normal ageing: A review of the efficiency of clinical and experimental memory measures," *Journal of clinical and experimental neuropsychology*, vol. 25, pp. 216–33, 05 2003.

[17] C. Trevisan, N. Veronese, S. Maggi, G. Baggio, E. Toffanello, S. Zambon, L. Sartori, E. Musacchio, E. Perissinotto, G. Crepaldi, E. Manzato, and G. Sergi, "Factors influencing transitions between frailty states in elderly adults: The progetto veneto anziani longitudinal study," *Journal of the American Geriatrics Society*, vol. 65, 11 2016.

[18] S. Gale, D. Acar, and K. Daffner, "Dementia," *The American Journal of Medicine*, vol. 131, 02 2018.

[19] S. Gillies *et al.*, "Shapely: manipulation and analysis of geometric objects," toblerity.org, 2007–. [Online]. Available: https://github.com/Toblerity/Shapely

[20] M. Sturman, D. Vaillancourt, and D. Corcos, "Effects of aging on the regularity of physiological tremor," *Journal of neurophysiology*, vol. 93, pp. 3064–74, 07 2005.

[21] F. Pasquier, "Early diagnosis of dementia: Neuropsychology," *Journal of neurology*, vol. 246, pp. 6–15, 02 1999.

[22] A. Connolly, E. Gaehl, H. Martin, J. Morris, and N. Purandare, "Underdiagnosis of dementia in primary care: Variations in the observed prevalence and comparisons to the expected prevalence," *Aging and mental health*, vol. 15, pp. 978–84, 07 2011.

[23] R. Saxon, M. Gray, and O. Florin, "Reducing geriatric outpatient waiting times: Impact of an advanced health practitioner," *Australasian Journal on Ageing*, vol. 37, 10 2017.

[24] L. Lang, A. Clifford, L. Wei, D. Zhang, D. Leung, G. Augustine, I. Danat, W. Zhou, J. Copeland, K. Anstey, and R. Chen, "Prevalence and determinants of undetected dementia in the community: A systematic literature review and a meta-analysis," *BMJ Open*, vol. 7, p. e011146, 02 2017.

[25] F. Lunardini, M. Luperto, K. Daniele, N. Basilico, S. Damanti, C. Abbate, D. Mari, M. Cesari, S. Ferrante, and A. Borghese, "Validity of digital trail making test and bells test in elderlies," 05 2019, pp. 1–4.

[26] P. Osterrieth, "Filetest de copie d'une figure complex: Contribution a l'etude de la perception et de la memoire [the test of copying a complex figure: A contribution to the study of perception and memory]," *Archives De Psychologie*, vol. 30, pp. 286–356, 01 1994.

[27] J. Liberman, W. Stewart, O. Seines, and B. Gordon, "Rater agreement for the rey-osterrieth complex figure test," *Journal of clinical psychology*, vol. 50, pp. 615–24, 08 1994.

[28] M. Concha, N. Graham, A. Munoz, D. Vlahov, W. Royal III, M. Updike, T. Nance-Sproson, O. Selnes, and J. Mcarthur, "Effect of chronic substance abuse on the neuropsychological performance of intravenous drug users with a high prevalence of hiv-1 seropositivity," *American journal of epidemiology*, vol. 136, pp. 1338–48, 01 1993.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," vol. 9908, 10 2016, pp. 630–645.

[30] J. Schneider, Z. Arvanitakis, W. Bang, and D. Bennett, "Schneider ja, arvanitakis z, bang w, bennett damixed brain pathologies account for most dementia cases in community-dwelling older persons. neurology 69:2197-2204," *Neurology*, vol. 69, pp. 2197–204, 03 2008.

[31] L. Lau, P. Giesbergen, M. Rijk, A. Hofman, P. Koudstaal, and M. Breteler, "Incidence of parkinsonism and parkinson disease in a general population: The rotterdam study," *Neurology*, vol. 63, pp. 1240–4, 11 2004.

[32] "World economic and social survey 2007: development in an ageing world," New York: United Nations Department of Social and Economic Affairs, Tech. Rep. E/2007/50/Rev.1 ST/ESA/314, 2007.

[33] K. Christensen, G. Doblhammer, R. Rau, and J. Vaupel, "Ageing populations: The challenges ahead," *Lancet*, vol. 374, pp. 1196–208, 10 2009.

[34] "Global population ageing: peril or promise?" Geneva: World Economic Forum, Tech. Rep. 4-13, 2012.

[35] E. Crimmins and H. BeltrÃ¡n-SÃ¡nchez, "Mortality and morbidity trends: Is there compression of morbidity?" *The journals of gerontology. Series B, Psychological sciences and social sciences*, vol. 66, pp. 75–86, 01 2011.

[36] K. Manton, X. Gu, and V. Lamb, "Change in chronic disability from 1982 to 2004/2005 as measured by long-term changes in function and health in the u.s. elderly population," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, pp. 18 374–9, 12 2006.

[37] S.-F. Lin, A. Beck, B. Finch, R. Hummer, and R. Masters, "Trends in us older adult disability: Exploring age, period, and cohort effects," *American journal of public health*, vol. 102, pp. 2157–63, 09 2012.

[38] "Ageing in ireland," Dublin: Central Statistics Office, Tech. Rep., 2007.

[39] S. Watanabe, S. Tsugane, T. Sobue, M. Konishi, and S. Baba, "Study design and organization of the jphc study. japan public health center-based prospective study on cancer and cardiovascular diseases," *Journal of epidemiology / Japan Epidemiological Association*, vol. 11, pp. S3–7, 11 2001.

[40] T. Kirkwood, "A systematic look at an old problem," *Nature*, vol. 451, pp. 644–7, 03 2008.

[41] A. Cruz-Jentoft, G. Bahat, J. Bauer, Y. Boirie, O. Bruyere, T. Cederholm, C. Cooper, F. Landi, Y. Rolland, A. Aihie Sayer, S. Schneider, C. Sieber, E. TopinkovÃ¡, M. Vandewoude, M. Visser, M. Zamboni, and A. Cherubini, "Sarcopenia: revised european consensus on definition and diagnosis," *Age and ageing*, vol. 48, 10 2018.

[42] T. Rantanen, S. Volpato, M. PhD, M. PhD, L. Fried, and P. MD, "Handgrip strength and cause-specific and total mortality in older disabled women: Exploring the mechanism," *Journal of the American Geriatrics Society*, vol. 51, pp. 636 – 641, 04 2003.

[43] J. Compston, M. Mcclung, and W. Leslie, "Osteoporosis," *The Lancet*, vol. 393, pp. 364–376, 01 2019.

[44] C. Novelli, J. Costa, and R. Souza, "Effects of aging and physical activity on articular cartilage: A literature review," *Journal of Morphological Sciences*, vol. 29, pp. 1–7, 01 2012.

[45] S. Studenski, S. Perera, K. Patel, C. Rosano, K. Faulkner, M. Inzitari, J. Brach, J. Chandler, P. Cawthon, E. Connor, M. Nevitt, M. Visser, S. Kritchevsky, S. Badinelli, T. Harris, A. Newman, J. Cauley, L. Ferrucci, and J. Guralnik, "Gait speed and survival in older adults," *JAMA : the journal of the American Medical Association*, vol. 305, pp. 50–8, 01 2011.

[46] F. Sorond, Y. Cruz-Almeida, D. Clark, A. Viswanathan, C. Scherzer, P. De Jager, A. Csiszar, P. Laurienti, J. Hausdorff, W. Chen, L. Ferrucci, C. Rosano, S. Studenski, S. Black, and L. Lipsitz, "Aging, the central nervous system, and mobility in older adults: Neural mechanisms of mobility impairment," *The journals of gerontology. Series A, Biological sciences and medical sciences*, vol. 70, 09 2015.

[47] A. Hickenbotham, A. Roorda, C. Steinmaus, and A. Glasser, "Meta-analysis of sex differences in presbyopia," *Investigative ophthalmology and visual science*, vol. 53, pp. 3215–20, 04 2012.

[48] K. Parham, B. McKinnon, D. Eibling, and G. Gates, "Challenges and opportunities in presbycusis," *Otolaryngology-head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery*, vol. 144, pp. 491–5, 04 2011.

[49] E. Glisky, *Changes in Cognitive Function in Human Aging*, 04 2007.

[50] T. Zanto and A. Gazzaley, "Attention and ageing," *The Oxford Handbook of Attention*, pp. 927–971, 01 2014.

[51] J. Henry, M. MacLeod, L. Phillips, and J. Crawford, "A meta-analytic review of prospective memory and aging," *Psychology and aging*, vol. 19, pp. 27–39, 04 2004.

[52] A. Muscari, C. Giannoni, L. Pierpaoli, A. Berzigotti, P. Maietta, E. Foschi, C. Ravaioli, G. Poggiopollini, G. Bianchi, D. Magalotti, C. Tentoni, and M. Zoli, "Chronic endurance exercise training prevents

aging-related cognitive decline in healthy older adults: A randomized controlled trial," *International journal of geriatric psychiatry*, vol. 25, pp. 1055–64, 10 2009.

[53] R. Roberts, "Mci incidence, progression to dementia, and reversion to normal in a population-based cohort: The mayo clinic study of aging," *Alzheimers and Dementia - ALZHEIMERS DEMENT*, vol. 7, 07 2011.

[54] L. Ginesi, C. Jenkins, and B. Keenan, "How dementia differs from normal ageing," vol. 112, pp. 12–15, 06 2016.

[55] C. Steves, T. Spector, and S. Jackson, "Ageing, genes, environment and epigenetics: What twin studies tell us now, and in the future," *Age and ageing*, vol. 41, pp. 581–6, 07 2012.

[56] A. Marengoni, S. Angleman, R. Melis, F. Mangialasche, A. Karp, A. Garmen, B. Meinow, and L. Fratiglioni, "Aging with multimorbidity: A systematic review of the literature," *Ageing research reviews*, vol. 10, pp. 430–9, 03 2011.

[57] A. Batko-Szwaczka, K. Wilczynski, B. Hornik, M. Janusz-Jenczen, I. Wlodarczyk, B. Wnuk, J. Szoltysek, J. Durmala, K. Szuster-Kowolik, K. Antoniak-Sobczak, J. Dulawa, and J. Szewieczek, "Predicting adverse outcomes in healthy aging community-dwelling early-old adults with the timed up and go test," *Clinical Interventions in Aging*, vol. Volume 15, pp. 1263–1270, 07 2020.

[58] M. Cesari, M. Prince, A. Jotheeswaran, I. Araujo de Carvalho, R. Bernabei, P. Chan, L. Gutierrez-Robledo, j.-p. Michel, J. Morley, P. Ong, L. Rodriguez-Manas, A. Sinclair, C. Won, J. Beard, and B. Vellas, "Frailty: An emerging public health priority," *Journal of the American Medical Directors Association*, vol. 17, 01 2016.

[59] B. Santos-Eggimann, P. Cuenoud, J. Spagnoli, and J. Junod, "Prevalence of frailty in middle-aged and older community-dwelling europeans living in 10 countries," *The journals of gerontology. Series A, Biological sciences and medical sciences*, vol. 64, pp. 675–81, 04 2009.

[60] M. Markle-Reid and G. Browne, "Conceptualizations of frailty in relation to older adults," *Journal of advanced nursing*, vol. 44, pp. 58–68, 11 2003.

[61] T. Gill, E. Gahbauer, H. Allore, and L. Han, "Transitions between frailty states among community-living older persons," *Archives of internal medicine*, vol. 166, pp. 418–23, 03 2006.

[62] R. Daniels, E. Rossum, L. Witte, G. Kempen, and W. Heuvel, "Interventions to prevent disability in frail community-dwelling elderly: A systematic review," *BMC health services research*, vol. 8, p. 278, 01 2009.

[63] A. Mitnitski, A. Mogilner, and K. Rockwood, "Accumulation of deficits as a proxy measure of aging," *TheScientificWorldJournal*, vol. 1, pp. 323–36, 09 2001.

[64] K. Rockwood and A. Mitnitski, "Frailty in relation to the accumulation of deficits," *The journals of gerontology. Series A, Biological sciences and medical sciences*, vol. 62, pp. 722–7, 08 2007.

[65] E. Hoogendijk, J. Afilalo, K. Ensrud, P. Kowal, G. Onder, and L. Fried, "Frailty: implications for clinical practice and public health," *The Lancet*, vol. 394, pp. 1365–1375, 10 2019.

[66] N. Peel, H. Bartlett, and R. McClure, "Healthy ageing: How is it defined and measured?" *Australasian Journal on Ageing*, vol. 23, 09 2004.

[67] R. J. Beard, A. Jotheeswaran, M. Cesari, and I. Araujo de Carvalho, "The structure and predictive value of intrinsic capacity in a longitudinal study of ageing," *BMJ Open*, vol. 9, 2019.

[68] C. Black, H. Fillit, L. Xie, X. Hu, M. Kariburyo, B. Ambegaonkar, O. Baser, H. Yuce, and R. Khandker, "Economic burden, mortality, and institutionalization in patients newly diagnosed with alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 61, pp. 1–9, 10 2017.

[69] M. Kabiri, M. Brauer, J. Shafrin, J. Sullivan, T. Gill, and D. Goldman, "Long-term health and economic value of improved mobility among older adults in the united states," *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, vol. 21, pp. 792–798, 07 2018.

[70] J. Frenk, L. Chen, Z. Bhutta, J. Cohen, N. Crisp, T. Evans, H. Fineberg, P. Garcia, Y. Ke, P. Kelley, B. Kistnasamy, A. Meleis, D. Naylor, A. Pablos-Mendez, S. Reddy, S. Scrimshaw, J. SepÃºlveda, D. Serwadda, and H. Zurayk, "Health professionals for a new century: Transforming education to strengthen health systems in an interdependent world," *Lancet*, vol. 376, pp. 1923–58, 12 2010.

[71] S. Pruitt and J. Epping-Jordan, "Preparing the 21st century global healthcare workforce," *BMJ (Clinical research ed.)*, vol. 330, pp. 637–9, 04 2005.

[72] L.-F. Low, M. Yap, and H. Brodaty, "A systematic review of different models of home and community care services for older persons," *BMC health services research*, vol. 11, p. 93, 05 2011.

[73] C. Ham, "The ten characteristics of high-performing chronic care system," *Health economics, policy, and law*, vol. 5, pp. 71–90, 10 2009.

[74] "Health at a glance: Europe 2018: State of health in the eu cycle," Brussels: OECD Publishing, Paris/European Union, Tech. Rep. https://doi.org/10.1797/health-glance-eur-2018-en., 2018.

[75] A. Wong, P. Baal, H. Boshuizen, and J. Polder, "Exploring the influence of proximity to death on disease-specific hospital expenditures: A carpaccio of red herrings," *Health economics*, vol. 20, pp. 379–400, 04 2011.

[76] P. Marik, "The cost of inappropriate care at the end of life: Implications for an aging population," *The American journal of hospice and palliative care*, vol. 32, 06 2014.

[77] T. Fulop, A. Larbi, J. Witkowski, J. McElhaney, M. Loeb, A. Mitnitski, and G. Pawelec, "Aging, frailty and age-related disease," *Biogerontology*, vol. 11, pp. 547–63, 10 2010.

[78] A. M. Beck, S. Kjaer, B. Hansen, R. Storm, K. Thal-Jantzen, and C. Bitz, "Follow-up home visits with registered dietitians have a positive effect on the functional and nutritional status of geriatric medical patients after discharge: A randomized controlled trial," *Clinical rehabilitation*, vol. 27, 12 2012.

[79] R. Moynihan, S. Sanders, Z. Michaleff, A. Scott, J. Clark, E. To, M. Jones, E. Kitchener, M. Fox, M. Johansson, E. Lang, A. Duggan, I. Scott, and L. Albarqouni, "Pandemic impacts on healthcare utilisation: a systematic review," 10 2020.

[80] "Ama digital health research: Physician's motivation and requirements for adopting digital health-adoption and attitudinal shifts from 2016 to 2019," American Medical Association, Tech. Rep. https://www.ama-assn.org/system/files/2020-02/ama-digital-health-study.pdf., 2020.

[81] K. Martinez, M. Rood, N. Jhangiani, L. Kou, S. Rose, A. Boissy, and M. Rothberg, "Patterns of use and correlates of patient satisfaction with a large nationwide direct to consumer telemedicine service," *Journal of General Internal Medicine*, vol. 33, 08 2018.

[82] R. Bashshur, C. Doarn, J. Frenk, J. Kvedar, and J. Woolliscroft, "Telemedicine and the covid-19 pandemic, lessons for the future," *Telemedicine and e-Health*, vol. 26, 04 2020.

[83] M. O'Connor, U. As, M. Dempsey, A. Huffenberger, S. Jost, D. Flynn, and A. Norris, "Using telehealth to reduce all-cause 30-day hospital readmissions among heart failure patients receiving skilled home health services," *Applied Clinical Informatics*, vol. 7, pp. 238–247, 04 2016.

[84] O. Taiwo and A. Ezugwu, "Smart healthcare support for remote patient monitoring during covid-19 quarantine," *Informatics in Medicine Unlocked*, vol. 20, 09 2020.

[85] W. Sulistyawati, I. Jayani, S. Susmiati, E. Yunalia, and A. Etika, "Use of telehealth in outpatients settings during the pandemic covid-19: A literature review," *Journal for Quality in Public Health*, vol. 4, pp. 157–160, 11 2020.

[86] R. Evering, J. Makonga-Braaksma, M. Uitdehaag, and M. Ouden, "Telehealth," *Pallium*, vol. 22, pp. 23–25, 11 2020.

[87] "Active ageing: a policy framework," Geneva: World Health Organization, Tech. Rep. WHO/NMH/NPH/02.8, 2002.

[88] N. Goodwin, L. Sonola, V. Thiel, and D. Kodner, "Co-ordinated care for people with complex chronic conditions: key lessons and markers for success," London: The King's Fund, Tech. Rep., 2013.

[89] K. Mcdonald, E. Schultz, and C. Chang, "Evaluating the state of quality-improvement science through evidence synthesis: Insights from the closing the quality gap series," *The Permanente journal*, vol. 17, 09 2013.

[90] M. Ouwens, H. Wollersheim, R. Hermens, M. Hulscher, and R. Grol, "Integrated care programmes for chronically ill patients: A review of systematic reviews," *International journal for quality in health care : journal of the International Society for Quality in Health Care / ISQua*, vol. 17, pp. 141–6, 05 2005.

[91] "Who global strategy on people-centred and integrated health services," Geneva: World Health Organization, Tech. Rep. http://www.who.int/servicedeliverysafety/areas/people-centred-care/global-strategy/en/, 2015.

[92] T. Perry, T. Andersen, and D. Kaplan, "Relocation remembered: Perspectives on senior transitions in the living environment," *The Gerontologist*, vol. 54, 07 2013.

[93] J. Wiles, A. Leibing, N. Guberman, J. Reeve, and R. Allen, "The meaning of "aging in place" to older people," *The Gerontologist*, vol. 52, pp. 357–66, 10 2011.

[94] H. Chaudhury, H. Cooke, H. Cowie, and L. Razaghi, "The influence of the physical environment on residents with dementia in long-term care settings: A review of the empirical literature," *The Gerontologist*, vol. 58, 03 2017.

[95] K. Marek, F. Stetzer, S. Adams, L. Popejoy, and M. Rantz, "Aging in place versus nursing home care comparison of costs to medicare and medicaid," *Research in gerontological nursing*, vol. 5, pp. 123–9, 08 2011.

[96] S. Golant, "Commentary: Irrational exuberance for the aging in place of vulnerable low-income older homeowners," *Journal of aging and social policy*, vol. 20, pp. 379–97, 02 2008.

[97] M. Pashmdarfard and A. Azad, "Assessment tools to evaluate activities of daily living (adl) and instrumental activities of daily living (iadl) in older adults: A systematic review," *Medical journal of the Islamic Republic of Iran*, vol. 34, pp. 224–239, 04 2020.

[98] H. Guo and A. Sapra, *Instrumental Activity of Daily Living (IADL) - PMID: 31985920*, 12 2019.

[99] "Who world health survey. in: World health organization, health statistics and information systems," Geneva: World Health Organization, Tech. Rep. http://www.who.int/healthinfo/survey/en/, accessed 23 June 2015.

[100] D. Elvira and O. Mascarilla, "'ageing in place'? exploring elderly people's housing preferences in spain," *Urban Studies*, vol. 46, pp. 295–316, 02 2009.

[101] "ehealth and ageing," Brussels: European Commission, Tech. Rep. http://ec.europa.eu/digital-agenda/en/ehealth-and-ageing, 2015.

[102] D. Misra, G. Das, and D. Das, *Review on Internet of Things (IoT): Making the World Smart*, 01 2018, pp. 827–836.

[103] V. Ricquebourg, D. Menga, D. Durand, B. Marhic, and C. Loge, "The smart home concept : our immediate future," 01 2007, pp. 23 – 28.

[104] D. Pal, T. Triyason, and S. Funilkul, "Smart homes and quality of life for the elderly: A systematic review," 12 2017, pp. 413–419.

[105] D. Febbo, F. Lunardini, M. Malavolti, A. Pedrocchi, A. Borghese, and S. Ferrante, "Iot ink pen for ecological monitoring of daily life handwriting," vol. 2020, 07 2020, pp. 5749–5752.

[106] L. Gemert-Pijnen, *eHealth Research, Theory and Development: A Multi-Disciplinary Approach*, 05 2018.

[107] C. Spinuzzi, "The methodology of participatory design," *Technical Communication*, vol. 52, pp. 163–174, 05 2005.

[108] D. Southwick, G. Resch, and M. Ratto, *Iterative Prototyping and Co-design*, 01 2021, pp. 231–238.

[109] J. Sturm, *Persuasive Technology*, 08 2017.

[110] K. Patrick, E. Hekler, D. Estrin, D. Mohr, H. Riper, D. Crane, J. Godino, and W. Riley, "The pace of technologic change," *American Journal of Preventive Medicine*, vol. 51, pp. 816–824, 11 2016.

[111] B. Wright, C. Carver, and M. Scheier, "On the self-regulation of behavior," *Contemporary Sociology*, vol. 29, p. 386, 03 2000.

[112] H. Kip, S. Kelders, Y. Bouman, and L. Gemert-Pijnen, "The importance of systematically reporting and reflecting on ehealth development: Participatory development process of a virtual reality application for forensic mental health care," *Journal of Medical Internet Research*, vol. 21, p. e12972, 08 2019.

[113] J. Sjostrom, L. Essen, and H. Gronqvist, "The origin and impact of ideals in ehealth research: Experiences from the u-care research environment," *JMIR research protocols*, vol. 3, p. e28, 05 2014.

[114] M. Naderifar, H. Goli, and F. Ghaljaei, "Snowball sampling: A purposeful method of sampling in qualitative research," *Strides in Development of Medical Education*, vol. In Press, 09 2017.

[115] N. Beerlage-de Jong, A. Eikelenboom-Boskamp, A. Voss, R. Sanderman, and J. Gemert-Pijnen, "Combining user-centered design with the persuasive systems design model: The development process of a web-based registration and monitoring system for healthcare-associated infections in nursing homes," *International Journal on Advances in Life Sciences*, vol. 6, pp. 262–271, 01 2014.

[116] A. Foulonneau, G. Calvary, and E. Villain, "State of the art in persuasive systems design," 06 2015.

[117] B. Fogg, "A behavior model for persuasive design," 01 2009, p. 40.

[118] M. Jaspers, "A comparison of usability methods for testing interactive health technologies: Methodological aspects and empirical evidence," *International journal of medical informatics*, vol. 78, pp. 340–53, 12 2008.

[119] D. Marks, *Randomized controlled trials*, 10 2020.

[120] R. Crosby, J. Lavender, S. Engel, and S. Wonderlich, *Ecological Momentary Assessment*, 01 2017, pp. 317–320.

[121] M. Carter, *Health Technology Assessment*, 06 2020.

[122] H. Atlam and G. Wills, *IoT Security, Privacy, Safety and Ethics*, 03 2019, pp. 1–27.

[123] E. Brooks, C. Turvey, and E. Augusterfer, "Provider barriers to telemental health: Obstacles overcome, obstacles remaining," *Telemedicine journal and e-health : the official journal of the American Telemedicine Association*, vol. 19, 04 2013.

[124] J. Augusto Wrede and P. Mccullagh, "Ambient intelligence: Concepts and applications," *Comput. Sci. Inf. Syst.*, vol. 4, pp. 1–27, 01 2007.

[125] S. Planton, M. Jucla, F.-E. Roux, and J. Demonet, "The 'handwriting brain': A meta-analysis of neuroimaging studies of motor versus orthographic processes," *Cortex; a journal devoted to the study of the nervous system and behavior*, vol. 49, 06 2013.

[126] M. Caligiuri, C. Snell, S. Park, and J. Corey-Bloom, "Handwriting movement abnormalities in symptomatic and premanifest huntington's disease," *Movement Disorders Clinical Practice*, vol. 6, 08 2019.

[127] J. Alty, J. Cosgrove, D. Thorpe, and P. Kempster, "How to use pen and paper tasks to aid tremor diagnosis in the clinic," *Practical Neurology*, vol. 17, pp. practneurol–2017, 08 2017.

[128] J. Kawa, A. Bednorz, P. Stepien, J. Derejczyk, and M. Bugdol, "Spatial and dynamical handwriting analysis in mild cognitive impairment," *Computers in Biology and Medicine*, vol. 82, 01 2017.

[129] S. Rosenblum, B. Engel-Yeger, and Y. Fogel, "Age-related changes in executive control and their relationships with activity performance in handwriting," *Human movement science*, vol. 32, pp. 1056–69, 10 2013.

[130] R. Camicioli, S. Mizrahi, J. Spagnoli, C. Bula, J. Demonet, F. Vingerhoets, A. Gunten, and B. Santos-Eggimann, "Handwriting and prefrailty in the lausanne cohort 65+ (lc65+) study," *Archives of Gerontology and Geriatrics*, vol. 61, 01 2015.

[131] S. Gerth, A. Klassert, T. Dolk, M. Brenner-Fliesser, M. Fischer, G. Nottbusch, and J. Festman, "Is handwriting performance affected by the writing surface? comparing tablet vs. paper," 01 2016.

[132] S. Mukherjee, *Anomaly Detection*, 01 2021, pp. 145–158.

[133] I. Deary, J. Corley, A. Gow, S. Harris, L. Lopez, R. Marioni, L. Penke, S. Rafnsson, and J. Starr, "Age-associated cognitive decline," *British medical bulletin*, vol. 92, pp. 135–52, 09 2009.

[134] S. L. Hong, E. James, and K. Newell, "Coupling and irregularity in the aging motor system: Tremor and movement," *Neuroscience letters*, vol. 433, pp. 119–24, 04 2008.

[135] N. Drempt, A. Mccluskey, and N. Lannin, "Handwriting in healthy people aged 65years and over," *Australian occupational therapy journal*, vol. 58, pp. 276–86, 08 2011.

[136] G. Vessio, "Dynamic handwriting analysis for neurodegenerative disease assessment: A literary review," *Applied Sciences*, vol. 9, p. 4666, 11 2019.

[137] O. Tucha, L. Tucha, C. Smely, and K. Lange, "Effects of excision of a mass lesion of the precentral region of the left hemisphere on disturbances of graphomotor output," *Neuroscience and Medicine*, vol. 03, 01 2012.

[138] L. M.G. and H. R.A., "A nonlinear analysis of the temporal characteristics of handwriting," *Human Movement Science*, vol. 18, pp. 485–524, 1999.

[139] A. Meigal, S. Rissanen, M. Tarvainen, S. Georgiadis, P. Karjalainen, O. Airaksinen, and M. Kankaanpaa, "Linear and nonlinear tremor acceleration characteristics in patients with parkinson's disease," *Physiological measurement*, vol. 33, pp. 395–412, 03 2012.

[140] R. Edwards and A. Beuter, "Dynamic handwriting analysis for neurodegenerative disease assessment: A literary review," *IEEE Transactions on Biomedical Engineering*, vol. 46(7), pp. 895–898, 1999.

[141] J. Garre-Olmo, M. Faundez-Zanuy, K. Lopez-de Ipina, L. Calvo-Perxas, and O. Turro-Garriga, "Kinematic and pressure features of handwriting and drawing: Preliminary results between patients with mild cognitive impairment, alzheimer disease and healthy controls," *Current Alzheimer research*, vol. 14, 05 2016.

[142] F. Simona, P. Alessandra, L. Francesca, M. Milad, D. F. Davide, and B. Nunzio Alberto, "Strumento di scrittura, sistema e metodo per il monitoraggio e l'analisi trasparente della scrittura," Italian Patent Application n. 10202000000679, priority date 31/03/2020.

[143] R. Rajamani, N. Piyabongkarn, V. Tsourapas, and J. Lew, "Parameter and state estimation in vehicle roll dynamics," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, pp. 1558–1567, 12 2011.

[144] J. Bland and D. Altman, "Statistical-methods for assessing agreement between 2 methods of clinical measurement," *International journal of nursing studies*, vol. 47, 04 2010.

[145] P. Werner and A. Korczyn, "Willingness to use computerized systems for the diagnosis of dementia testing a theoretical model in an israeli sample," *Alzheimer disease and associated disorders*, vol. 26, pp. 171–8, 06 2011.

[146] T. Asselborn, P. Dillenbourg, and M. Chapatte, "Extending the spectrum of dysgraphia: A data driven strategy to estimate handwriting quality," *Scientific Reports*, vol. 10, 02 2020.

[147] G. Cascarano, C. Loconsole, A. Brunetti, A. Lattarulo, D. Buongiorno, G. Losavio, E. Di Sciascio, and V. Bevilacqua, "Biometric handwriting analysis to support parkinson's disease assessment and grading," *BMC Medical Informatics and Decision Making*, vol. 19, p. 252, 12 2019.

[148] Y.-L. Hsu, C.-L. Chu, Y.-J. Tsai, and J.-S. Wang, "An inertial pen with dynamic time warping recognizer for handwriting and gesture recognition," *Sensors Journal, IEEE*, vol. 15, pp. 154–163, 01 2015.

[149] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C.-C. Tung, and H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. A*, vol. 454, pp. 679–699, 01 1998.

[150] J. Zhang, B. Price, R. Adams, K. Burbank, and T. Knaga, "Detection of involuntary human hand motions using empirical mode decomposition and hilbert-huang transform," 09 2008, pp. 157 – 160.

[151] P. O'Suilleabhain and J. Matsumoto, "Time-frequency analysis of tremors," *Brain : a journal of neurology*, vol. 121 (Pt 11), pp. 2127–34, 12 1998.

[152] S. Pincus, "Approximate entropy as a measure of system complexity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, pp. 2297–301, 04 1991.

[153] D. Vaillancourt, A. Slifkin, and K. Newell, "Regularity of force tremor in parkinson's disease," *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, vol. 112, pp. 1594–603, 10 2001.

[154] N. Marwan, M. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics Reports*, vol. 438, pp. 237–329, 01 2007.

[155] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena*, vol. 110, pp. 43–50, 12 1997.

[156] D. Febbo, F. Lunardini, M. Malavolti, A. Pedrocchi, A. Borghese, and S. Ferrante, "Iot ink pen to study age-related changes in handwriting." GNB2020 - Seventh National Congress of Bioengineering - Proceedings 2020, 2020.

[157] T. Koo and M. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *Journal of Chiropractic Medicine*, vol. 15, 03 2016.

[158] H. De Vet, C. Terwee, R. Ostelo, H. Beckerman, D. Knol, and L. Bouter, "Minimal changes in health status questionnaires: Distinction between minimally detectable change and minimally important change," *Health and quality of life outcomes*, vol. 4, p. 54, 02 2006.

[159] C. Yu, "Test-retest reliability," *Encyclopedia of social measurement*, vol. 3, pp. 777–784, 01 2005.

[160] F. Lunardini, D. Di Febbo, and al..., "A smart ink pen for the ecological assessment of age-related changes in writing and tremor features," *IEEE Transactions on Instrumentation and Measurement*, 2020.

[161] B. Engel-Yeger, S. Hus, and S. Rosenblum, "Age effects on sensory-processing abilities and their impact on handwriting," *Canadian journal of occupational therapy. Revue canadienne d'ergotherapie*, vol. 79, pp. 264–74, 12 2012.

[162] A. Letanneux, D. J., J. Velay, F. Viallet, and S. Pinto, "From micrographia to parkinson's disease dysgraphia," *Mov. Disord.*, vol. 29, pp. 1467–1475, 10 2014.

[163] M. Thomas Kollamkulam, A. Lenka, and P. Pal, "Handwriting analysis in parkinson's disease: Current status and future directions," *Movement Disorders Clinical Practice*, vol. 4, 09 2017.

[164] M. Caligiuri, H.-L. Teulings, C. Dean, and J. Lohr, "A quantitative measure of handwriting dysfluency for assessing tardive dyskinesia," *Journal of clinical psychopharmacology*, vol. 35, 02 2015.

[165] A. Bisio, L. PedullÃ , L. Bonzano, A. Tacchino, G. Brichetto, and M. Bove, "The kinematics of handwriting movements as expression of cognitive and sensorimotor impairments in people with multiple sclerosis," *Scientific Reports*, vol. 7, 12 2017.

[166] L. G. Dui, F. Lunardini, C. Termine, M. Matteucci, and S. Ferrante, "A tablet-based app to discriminate children at potential risk of handwriting alterations in a preliteracy stage," vol. 2020, 07 2020, pp. 5856–5859.

[167] L. G. Dui, F. Lunardini, C. Termine, M. Matteucci, N. Stucchi, N. Borghese, and S. Ferrante, "A tablet app for handwriting skill screening at the preliteracy stage: Instrument validation study," *JMIR Serious Games*, vol. 8, p. e20126, 10 2020.

[168] F. Lunardini, C. Casellato, M. Bertucco, T. Sanger, and A. Pedrocchi, "Children with and without dystonia share common muscle synergies while performing writing tasks," *Annals of biomedical engineering*, vol. 45, 05 2017.

[169] R. Cohen, B. Cohen-Kroitoru, A. Halevy, S. Aharoni, I. Aizenberg, and A. Shuper, "Handwriting in children with attention deficient hyperactive disorder: role of graphology," *BMC Pediatrics*, vol. 19, 12 2019.

[170] K. Lakshmi, N. Neema, M. Muddasir, and M. Prashanth, *Anomaly Detection Techniques in Data Mining-A Review*, 01 2020, pp. 799–804.

[171] A. Larner, "Mini-mental state examination: diagnostic test accuracy study in primary care referrals," *Neurodegenerative Disease Management*, vol. 8, 09 2018.

[172] A. V. Dorogush, V. Ershov, and A. Gulin, "Catboost: gradient boosting with categorical features support," *arXiv preprint arXiv:1810.11363*, 2018.

[173] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in neural information processing systems*, 2017, pp. 4765–4774.

[174] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable ai for trees," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 2522–5839, 2020.

[175] A. E. Roth, *The Shapley value: essays in honor of Lloyd S. Shapley*. Cambridge University Press, 1988.

[176] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[177] G. Marzinotto, J. Rosales, M. El Yacoubi, S. Garcia-Salicetti, C. Kahindo, H. Kerherve', V. Cristancho-Lacroix, and A.-S. Rigaud, "Uncovering major age-related handwriting changes by unsupervised learning," 01 2016.

[178] G. Marzinotto, "Age-related evolution patterns in online handwriting," *Computational and Mathematical Methods in Medicine*, vol. 2016, 08 2016.

[179] S. Rosenblum, P. Werner, T. Dekel, I. Gurevitz, and J. Heinik, "Handwriting process variables among elderly people with mild major depressive disorder: A preliminary study," *Aging clinical and experimental research*, vol. 22, pp. 141–7, 04 2010.

[180] M. Baumgart, H. Snyder, M. Carrillo, S. Fazio, H. Kim, and H. Johns, "Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective," *Alzheimer's and dementia : the journal of the Alzheimer's Association*, vol. 10, 05 2015.

[181] S. Osburn, "The global prevalence of alzheimers: A systematic review and metaanalysis," *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde*, vol. 43, pp. 1307–12, 11 2015.

[182] S. Gale, D. Acar, and K. Daffner, "Dementia," *The American Journal of Medicine*, vol. 131, 02 2018.

[183] G. Roman and B. Pascual, "Contribution of neuroimaging to the diagnosis of alzheimer's disease and vascular dementia," *Archives of medical research*, vol. 43, 11 2012.

[184] J. Delrieu, A. Piau, C. Caillaud, T. Voisin, and B. Vellas, "Managing cognitive dysfunction through the continuum of alzheimer's disease," *CNS drugs*, vol. 25, pp. 213–26, 03 2011.

[185] M. Folstein, S. Folstein, and P. McHugh, *'Mini-Mental State'. A practical method for grading the cognitive state of patients for the clinician*, 12 1975, vol. 12.

[186] M. Berres, A. Monsch, F. Bernasconi, B. Thalmann, and H. Staehelin, "Normal ranges of neuropsychological tests for the diagnosis of alzheimer's disease," *Studies in health technology and informatics*, vol. 77, pp. 195–9, 02 2000.

[187] J. Meyers, *Trail Making Test*, 01 2017, pp. 1–2.

[188] L. Taylor, "Scoring criteria for the rey-osterrieth complex figure test," *A compendium of neuropsychological tests. Administration, norms, and commentary*, pp. 350–351, 01 1998.

[189] A. Rey and P. Osterrieth, "Translations of excerpts from andre rey's 'psychological examination of traumatic encephalopathy' and osterrieth's 'the complex figure test'," *Psychological Examination of Traumatic Encephalopathy and P. A. Osterrieth's the Complex Figure Test. Clinical Neuropsychologist*, vol. 7, pp. 2–21, 01 1993.

[190] E. Strauss, E. Sherman, and O. Spreen, "A compendium of neuropsychological tests: Administration, norms, and commentary," vol. 14, pp. 62–63, 01 2006.

[191] M. Grossman, S. Carvell, L. Peltzer, M. Stern, S. Gollomp, and H. Hurtig, "Visual construction impairments in parkinson's disease," *Neuropsychology*, vol. 7, pp. 536–547, 10 1993.

[192] C. Zucchella, A. Federico, A. Martini, M. Tinazzi, M. Bartolo, and S. Tamburin, "Neuropsychological testing," *Practical Neurology*, vol. 18, pp. practneurol–2017, 02 2018.

[193] L. Bertolani, E. Renzi, and P. Faglioni, "Test di memoria non verbale di impiego diagnostico in clinica: Taratura su soggetti normali," *Archivio di Psicologia, Neurologia e Psichiatria*, vol. 54, pp. 477–486, 01 1993.

[194] L. Tupler, K. Welsh, Y. Asare-aboagye, and D. Dawson, "Reliability of the rey-osterrieth complex figure in use with memory-impaired patients," *Journal of clinical and experimental neuropsychology*, vol. 17, pp. 566–79, 09 1995.

[195] W. Stewart, B. Gordon, O. Selnes, K. Bandeen-Roche, S. Zeger, R. Tusa, D. Celentano, A. Shechter, J. Liberman, and C. Hall, "A prospective study of cns function in the united states amateur boxers," *American journal of epidemiology*, vol. 139, pp. 573–88, 04 1994.

[196] R. Canham, S. Smith, and A. Tyrrell, "Automated scoring of a neuropsychological test: The rey osterrieth complex figure." vol. 2, 10 2000, pp. 2406–2413.

[197] J. Vogt, H. Kloosterman, S. Vermeent, G. van Elswijk, R. Dotsch, and B. Schmand, "Automated scoring of the rey-osterrieth complex figure test using a deep-learning algorithm," *Archives of Clinical Neuropsychology*, vol. 34, pp. 836–836, 08 2019.

[198] H. Levkowitz and G. Herman, "Glhs: A generalized lightness, hue, and saturation color model," *CVGIP: Graphical Models and Image Processing*, vol. 55, pp. 271–285, 07 1993.

[199] P. Rosin, "Unimodal thresholding," *Pattern Recognition*, vol. 34, pp. 2083–2096, 11 2001.

[200] S. Tambe, "Image processing (ip) through erosion and dilation methods," vol. Volume 3, pp. 285–289, 07 2008.

[201] W. Abuain, S. Abdullah, B. Bataineh, T. Abu-Ain, and K. Omar, "Skeletonization algorithm for binary images," *Procedia Technology*, vol. 11, 12 2013.

[202] R. Szeliski, *Computer Vision: Algorithms and Applications*, 01 2011, vol. 5.

[203] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[204] J. Matas, C. Galambos, and J. Kittler, "Robust detection of lines using the progressive probabilistic hough transform," *Computer Vision and Image Understanding*, vol. 78, pp. 119–137, 04 2000.

[205] S. Suzuki and K. be, "Topological structural analysis of digitized binary images by border following," *Computer Vision, Graphics, and Image Processing*, vol. 30, pp. 32–46, 03 1985.

[206] D. Douglas and T. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 10, pp. 112–122, 10 1973.

[207] G. Schmidtmann, B. Jennings, and F. Kingdom, "Shape recognition: Convexities, concavities and things in between," *Scientific Reports*, vol. 5, 11 2016.

[208] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *Proc. CVPR*, 03 2015.

[209] A. B. Jung, K. Wada, J. Crall, S. Tanaka, J. Graving, C. Reinders, S. Yadav, J. Banerjee, G. Vecsei, A. Kraft, Z. Rui, J. Borovec, C. Vallentin, S. Zhydenko, K. Pfeiffer, B. Cook, I. Fernandez, F.-M. De Rainville, C.-H. Weng, A. Ayala-Acevedo, R. Meudec, M. Laporte, *et al.*, "imgaug," https://github.com/aleju/imgaug, 2020, online; accessed 01-Feb-2020.

[210] T. Pereira and T. Campos, "Domain adaptation for person re-identification on new unlabeled data," 01 2020, pp. 695–703.

[211] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, 12 2014.

[212] A. Dorogush, V. Ershov, and A. Gulin, "Catboost: gradient boosting with categorical features support," 10 2018.

[213] T. Phan and K. Yamamoto, "Resolving class imbalance in object detection with weighted cross entropy losses," 06 2020.

[214] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," 12 2017.

[215] C. ji and S. Ma, "Combinations of weak classifiers," *Neural Networks, IEEE Transactions on*, vol. 8, pp. 32 – 42, 02 1997.