Towards frailty assessment in older adults: Investigation of sit-to-stand transfer detection using ear-worn sensors in real-world activities

Master's Thesis in Medical Engineering

submitted
by

Friederike Popp

born 07.04.1996 in Bayreuth

Written at

Machine Learning and Data Analytics Lab
Department Artificial Intelligence in Biomedical Engineering
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)

Advisors:
Ann-Kristin Seifer M. Sc., Robert Richer M. Sc., Prof. Dr. Bjoern Eskofier
(Machine Learning and Data Analytics Lab, FAU Erlangen-Nürnberg)
Dr. phil. Heiko Gaßner
(Department of Molecular Neurology, Universitiy Hospital Erlangen)

Started: 15.12.2022

Finished: 17.07.2023

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.
Die Richtlinien des Lehrstuhls für Bachelor- und Masterarbeiten habe ich gelesen und anerkannt, insbesondere die Regelung des Nutzungsrechts.

Erlangen, den 17.07.2023

Übersicht

Die Fähigkeit, Sit-to-Stand (STS)-Bewegungen auszuführen, zeigt den Gesundheitszustand und die Mobilität an. Herkömmliche Methoden zur Erfassung dieser Bewegungen sind teuer und auf Labore beschränkt. Tragbaren Sensoren wie Inertial Measurement Units (IMUs) bieten eine vielversprechende Methode zur Erkennung von STS-Bewegungen unter realen Bedingungen. Bisherige Studien konzentrierten sich auf Sensoren am unteren Rücken und an der Brust, aber nicht auf die vielversprechende Ohrposition. Weitere Studien zeigten, dass Frailty mittels STS-Bewegungen beurteilt werden kann. In dieser Arbeit wurde einen bestehenden Algorithmus zur STS-Erkennung mit IMUs am unteren Rücken, an der Brust und am Ohr zu validieren. Die Optimierung und Evaluierung des Algorithmus für diese Positionen führte zu vielen false positives (FPs), was die Anwendbarkeit unter realen Bedingungen einschränkt. Daher wurde ein zweiter Schritt entwickelt, bei dem zusätzliche Klassifikatoren für jede Position trainiert wurden. Nach der Optimierung zeigten die Brust- und Ohrpositionen eine bessere Performance (F1-Score 91.7 % bzw. 95.1 %), während der untere Rücken mit den Standardparametern einen höheren F1-Score von 97.3 % erreichte. Die Precision der Positionen am unteren Rücken und Ohr wurde durch den Machine Learning (ML)-Schritt auf 100 % verbessert. Ein weiteres Ziel dieser Arbeit ist die Beurteilung von Frailty andhand der erkannten STS-Bewegungen des zweistufigen Algorithmus. Es wurde eine Studie mit älteren, gebrechlichen Erwachsenen durchgeführt, die ein Hörgerät mit integrierter IMU trugen. Diese Daten wurde mit STS-Daten von älteren, gesunden Menschen kombiniert. Ein ML-Modell wurde trainiert, um den Frailty-Status zu bestimmen, indem Features aus den STS-Bewegungnen extrahiert wurden. Der Klassifikator erreichte einen F1-Score von 74.4 % mit wenig falsch zugeordneten Bewegungen und eine Genauigkeit von 73.5 %. Die Unterscheidung zwischen den frailen und nicht-frailen STS-Bewegungen wurde durch eine grapische Analyse der Features bestätigt. Diese Arbeit zeigt, dass der bestehende STS-Erkennungsalgorithmus durch die Verwendung eines zweistufigen Ansatzes an andere Sensorpositionen angepasst werden kann. Allerdings zeigte die Validierung an der Ohrposition bei älteren Probanden ergab jedoch eine schlechtere Performance mit vielen FPs und false negatives (FNs). Für die Vorhersage von Frailty erzielte der trainierte Klassifikator gute Ergebnisse, aber aufgrund des limitierten Datensatzes sollte man bei der Interpretation vorsichtig sein. Es handelt sich um einen theoretischen Ansatz zur Analyse der Machbarkeit, aber es wurde gezeigt, dass eine Unterscheidung zwischen frail und nicht-frail möglich ist.

Abstract

The ability to perform sit-to-stand (STS) transitions serves as an indicator of mobility and health. Traditional systems for assessing these movements are expensive and limited to laboratory settings. Wearable sensors, such as inertial measurement units (IMUs), offer a promising solution for real-world STS detection. Previous studies focused on sensors placed on the lower back and chest position and scripted STS transfers to detect transitions, but the ear position, which has shown potential in motion detection, was not investigated yet. Additionally, previous research had shown that STS transitions can be used to evaluate frailty. This thesis aims to validate an existing STS detection algorithm with IMUs placed on the lower back, chest, and ear. The algorithm was optimized and evaluated for these positions but yielded a high number of false positives (FPs) during daily tasks, limiting its applicability in real-world conditions. Therefore, a two-step approach was proposed consisting of the optimized hyperparameters for the algorithm and an additional trained classifier for each sensor position. After optimizing the algorithm, the chest and ear positions showed better performance (F1-score 91.7 % and 95.1 %, respectively), while the lower back position performed better with the default parameter set (F1-score 97.3 %). Applying the machine learning (ML)-step improved the precision of lower back and ear position to 100 %. Another aim of this thesis was to assess frailty status based on the detected STS transitions using the developed two-step approach. A study was conducted with older, frail adults wearing a hearing aid with an integrated IMU. This dataset was combined with another one containing STS data from elderly, healthy adults. A ML model was trained to predict the frailty status based on extracted features during STS transitions. The classifier achieved an F1-score of 74.4 % with few false classifications and accuracy of 73.5 % in correctly classified STS transitions. The graphical analysis further confirmed these results, indicating differences in features between the frail and non-frail STS transitions. This thesis demonstrates the adaptability of the existing STS detection algorithm to other sensor positions using a developed two-step approach. However, validation of this approach optimized for ear position in an older population showed poorer performance with a high number of FPs and false negatives (FNs). The classifier trained to predict frailty in older adults showed promising results, but caution is required due to the limitations of the data set, making it more of a theoretical approach to investigate feasibility. However, the ability to differentiate between the frail and non-frail groups was shown.

Contents

Chapter 1

Introduction

1.1 Motivation

Standing up from a chair is an essential activity during daily living [Nuz86]. The ability to perform these sit-to-stand (STS) transitions is important for maintaining independent living [Hug96], as it requires a combination of balance and muscle power [Atr20]. Therefore, a STS transition can provide valuable information about the functional mobility and overall health condition of an individual [Ada20a]. Functional performance tests involving STS transitions have proven to be effective tools for evaluating mobility impairments [Mil13a], frailty [Sav13], and fall risk in elderly adults [Cam89].

Frailty is a multidimensional geriatric syndrome [Ben21], describing the deficits developed during the aging process [Roc11]. Frailty in older adults is associated with an increasing risk for falls and hospitalization [Lac12]. The STS transitions can be used to assess the frailty status of individuals. In a clinical setting, the frailty status is typically assessed using questionnaires and functional performance tests. Due to the degenerative nature of frailty, short and episodic clinical interventions, and tests may not adequately assess this process. However, early detection is important to initiate appropriate interventions and to slow down the progression [Gre14]. Unobtrusive long-term measurements could offer objective information about disease progression.

Clinical assessments commonly use STS transitions to evaluate disease progression and the effectiveness of interventions [Ada20a]. To obtain more detailed information about the health status of an individual, motion capture systems [Abd21] or force plates have commonly been used for assessing STS movements [Ber04; Khe99; Mou00; Pai94]. However, these systems have disadvantages as they are restricted to laboratory environments and are expensive [Abd21; Sei12]. Additionally, clinic-based mobility assessments have limitations as they can only be performed episodically, therefore, are not able to map the whole range of motion of an individual [Ada20a]. Moreover, the laboratory settings differ from the natural environment, which limits the ability of clinical tests to accurately reflect natural performance [Kia97].

To overcome these limitations, wearable sensors, particularly inertial measurement units (IMUs), enable continuous monitoring of movements in both clinical and unsupervised real-world settings providing objective results within functional performance tests [Ada20a]. IMUs are small, lightweight sensors that enable continuous and unobtrusive measurements of movements by capturing three-dimensional acceleration and gyroscope data.

With this kind of sensors, long-term measurements are enabled. However, the disadvantage of long-term measurements, in general, is that there has to be worn a sensor in daily living. An important requirement of this sensor is therefore to be ubiquitous and non-intrusive. Fulfilling these requirements, the sensor can be seamlessly integrated into the life of an individual, especially for health monitoring of older people [Cob20]. Possible sensor positions for this application can be the head or ear position as a high number of individuals already wear devices on their ear during their daily-life activities, such as headphones or hearing aids [Röd22]. Therefore, exploring the potential of ear-worn IMUs in assessing movements such as STS transitions becomes interesting, especially for use in performance tests.

1.2 Related work

1.2.1 IMU-based automated STS detection

Several studies have investigated the potential of using IMUs for automated STS detection. Nguyen et al. [Ngu15] proposed a signal processing algorithm for automated recognition of activities of daily living based on multiple IMUs worn on different body segments. The study found the algorithm to be able to detect the activities of daily living from the extracted body segments from a set of IMUs with 100 % sensitivity and specificity, respectively.

In contrast, other studies have focused on using a single IMU. Some of these studies explored detecting STS transitions using multiple sensor modalities. An approach that investigated accelerometer data in combination with an air pressure sensor integrated into an IMU using a support vector machine (SVM) with a Gaussian kernel function within the detection algorithm was proposed by Zhang et al. [Zha14]. Various movement characteristics such as cross-correlation, intensity, orientation, and altitude change were used to train this algorithm for STS detection. With this algorithm an F1-score of 86 % was achieved in detecting STS transitions of older adults.

Another study used the two sensor modalities acceleration and angular velocity for assessing STS transitions in older adults using a single IMU on the lower back [van18]. They developed a signal processing algorithm for trunk angle calculation which was used for detecting STS transitions, providing detailed insights into the quality of movements. Lepetit et al. [Lep18] proposed a study using the same two sensor modalities, developing a motion detection algorithm to detect STS transitions. The signal-processing algorithm identified the start and end of a transition using thresholds for both signals determined within the data of this study. The kinetic energy during the transitions was calculated using the detected duration and this was used as a measurement indicator for assessing the health status.

Moreover, other studies investigated the detection of STS transitions using a single sensor modality from the IMU. This configuration proves to be more useful for long-term measurements in daily life. Different positions for the single sensor setup have been employed in these studies. A single IMU placed on the chest was used by Ganea et al. [Gan11] to detect STS transitions based on changes in the trunk tilt using only the gyroscope signal. The transition detection was done with an algorithm proposed by Salrian et al., originally developed for detecting tremors in Parkinson's disease (PD) patients [Sal07].

Another often used position was the lower back position. Millor et al. [Mil13a] investigated the use of a single IMU on the lower back for evaluating the 30 seconds chair stand test (30s CST) in an older prefrail population. The vertical position was calculated based on the vertical acceleration of the IMU for automatic detection of the STS cycles. Additionally, they determined three phases of the STS transition during the previously detected cycles. The study concluded that the introduced method was able to analyze the STS cycles within the 30s CST from IMU data automatically.

Hellmers et al. [Hel19] proposed another study for automated detection of STS transitions using the lower back position. The vertical acceleration was used for training and evaluating classifiers to get the different movements performed in the five times sit-to-stand (5xSTS) test. The multilayer perceptrons approach achieved an F1-score of 94.8 % for detecting the transitions. In a third study, a single IMU on the lower back was used to propose a signal-processing algorithm for detecting STS transitions [Ada20a]. The vertical acceleration was used for detecting the transitions using a Butterworth filter and continuous wavelet transform (CWT). This algorithm was used in this thesis for detecting the STS transitions.

Comparing the two sensor positions lower back and chest, Regterschot et al. [Reg15] proposed a method for automatic STS detection using a Butterworth filter on the vertical acceleration of the IMU sensors validated on chest position and the right side of the hip. This study revealed that the sensor on the chest position showed stronger associations with clinical measurements than the hip position.

Another investigated position for IMU placement in the detection of STS transitions was the thigh. A study explored this position for detecting the STS and stand-to-sit transitions using the peak angular velocity [Pic19]. Martinez-Hernandez et al. [Mar19] did an approach with a tigh placed IMU using the acceleration as a single sensor modality. The algorithm is based on Bayesian formulation and a sequential analysis method for identifying STS transitions. Sit, transitions, and stand activities were successfully identified by the proposed method and the three phases of a STS transition were recognized by an accuracy of 100 %.

To remove the dependency on specific sensor positions, Atrsaei et al. [Atr20] developed a location-independent algorithm for detecting postural transitions. The vertical acceleration in the global frame was used as a sensor modality for assessing the transitions comparing different sensor positions on the trunk. A CWT was used for getting potential transitions, validation of these transitions was done by using the vertical velocity of the IMU data. The best results were achieved for the sensor around the waist with a sensitivity over 90 % and a precision over 97 %. In general, the positions on the lower back and chest were found to be the most accurate for detecting postural transitions.

In summary, various studies have examined the detection of STS transitions using IMUs. The majority of these studies focused on sensors placed on the trunk, such as the lower back or chest. A few studies used the extremities, such as the tight, for sensor placement. No studies have investigated the detection of STS transitions using a sensor worn on the head or the ear.

1.2.2 Movement detection with ear-worn IMUs

IMUs worn on the ear position could be interesting for long-term measurements for assessing mobility and health status, as many people wear already devices on the ear, such as earphones or hearing aids, during daily living. Ear-worn IMUs are portable and can be worn for extended periods during the day due to their small and lightweight characteristics. These *earables* were among others used for gait detection, activity classification, and recognition, as well as for activity type recognition [Röd22].

Several studies investigated the use of earables for activity recognition tasks. Atallah et al. [Ata11] compared an ear-worn IMU sensor with six other sensor positions placed on the whole body during different activities of daily living including preparing food, getting dressed, and cleaning. The sensor attached to the knee was the best sensor for recognizing these activities and the ear position was second best, outperforming for example chest and waist position. Another study got similar results for comparing seven different sensor positions to classify different activities, such as climbing stairs, jumping, lying down, standing, sitting, running, and walking [Szt16]. Only the knee and shin performed better than the ear position, but the study also concluded that the sensors performed differently depending on the activity. Therefore, for different activities different sensor positions worked best.

Min et al. [Min18] applied a classifier on different activities, such as walking, stepping up, and down, with IMU sensors integrated into earbuds achieving a mean accuracy of 88 % for classifying the activities correctly. Similarly, Burgos et al. [Bur20] used ear-worn IMU sensors for classifying different gait activity tasks, which were walking and running. The study compared SVM and k-nearest neighbour (k-NN) as classifiers resulting in the same accuracy of over 99 % for both.

A comparison of the performance of a head-mounted IMU with a chest-mounted IMU for detecting postural transitions was done by Abdollah et al. [Abd21]. Within the study standardized performance tests including STS transitions were performed. The two sensor positions were compared to a motion capture system. An algorithm for the detection of postural transitions via head inclination angle and accelerometer was developed. As a result, the head-mounted device showed high accuracy, sensitivity, and specificity for detecting postural transitions and walking events compared to the chest-worn sensor.

In summary, earables find applications in a wide range of fields. They are capable of detecting the physiological state within health monitoring purposes, facilitating hands-free and eyes-free interaction with devices, and performing motion analysis tasks such as gait detection, activity classification, and activity recognition [Röd22]. The mentioned studies demonstrate that an IMU placed on the ear yields good results in these different tasks compared to other positions on the body.

1.2.3 Evaluation of frailty based on STS transitions

Different studies investigated the assessment of frailty in older people through STS transitions. Millor et al. [Mil13b] proposed a study in which the standardized 30s CST was used for assessing the transitions for evaluating the frailty status. They applied a previously deployed algorithm by Millor et al. [Mil13a] for automatic STS detection. In this study, kinematic parameters were defined for assessing the frailty level of an individual and it was found that the positive vertical velocity peaks during standing up and sitting down phases, and the area under curve (AUC) for the vertical acceleration were promising parameters for frailty classification within STS transitions [Mil13b].

Based on this study, Millor et al. [Mil17] proposed another study for investigating frailty status with IMU measurements. They used kinematic parameters from the 30s CST and the timed up and go (TUG) tests with a single IMU placed on the lower back. The previously proposed method for detecting STS transitions [Mil13a] and a decision tree analysis were employed for getting relevant frailty parameters and validating the classification. The selected parameters were found to be better in getting frailty status than the number of cycles performed and the normally used criteria including gait speed. Furthermore, it was found that with an increasing frailty level, the movement became slower, the trunk movement became greater, and there was a decrease in the vertical acceleration, velocity, and power peaks.

Another study explored the automatic and quantitative assessment of frailty status based on IMU parameters obtained during TUG performance with a shank-mounted IMU [Gre14]. A regression-based model was used for differentiating the two groups of frail and non-frail, achieving a mean classification accuracy of 72.8 %.

Besides these standardized performance tests, a study investigated the sensor-based assessment of functional status using normal STS transitions [Reg15]. The two sensor positions chest and lower back were compared for getting the frailty status. The chest sensor was found to be able to indicate the functional status of an individual, including frailty.

Zhang et al. [Zha17] investigated the STS performance through continuous measurements of daily living. Peak power measurements in daily life showed negative associations with clinical assessment of mobility, limitation in activities, and frailty. The results showed that the functional status of individuals could be assessed through long-term at-home measurements.

Another study, which investigated the frailty status in older people through daily postural transitions, was conducted by Parvaneh et al. [Par17]. The aim was to differentiate between non-frail and pre-frail/frail. The participants wore the IMU sensor for 48 hours, recording their daily postural activities, and an algorithm previously developed by Najafi et al. [Naj02] was used to identify the different postures. Additionally, an algorithm was developed to identify and count the number of postural transitions, as well as to calculate the ratio of cautious sitting. A lower total number of postural transitions per day and a higher ratio of cautious sitting were found to indicate frailty, although the accuracy of these indicators was not as high as that of the Fried frailty index.

Letpetit et al. [Lep19] conducted a study to develop a diagnostic tool for detecting functional deficits during STS transitions using a single IMU placed at the chest. The objective was to differentiate between frail older individuals and healthy older individuals, as well as healthy older individuals and young subjects. Frail individuals were classified as those with a frailty index greater than 5 according to the Rockwood index, which categorizes individuals into seven groups ranging from very fit (category 1) to severely frail (category 7) [Roc05]. A frailty score was established by performing a multifactorial analysis of various parameters, which were then reduced into a quantitative score using the first principal component of a principal component analysis (PCA). The maximal value of trunk velocity was significantly influenced by age and frailty, but a single parameter was deemed insufficiently accurate for evaluating STS performance [Lep19].

1.3 Purpose

Several prior studies have primarily focused on using IMUs placed on the lower back and chest for STS detection, showing consistent accuracy results in these positions. However, the ear position has shown promising outcomes in activity recognition tasks and motion detection, outperforming other positions. Yet, no investigations have been conducted on exploring the use of ear-worn IMUs for detecting STS transitions. Therefore, one aim of this thesis was to validate the performance of ear-worn IMUs in comparison to the lower back and chest position for STS transition detection. Additionally, another aim of this thesis was to use the features of STS transitions obtained through an algorithm using an ear-worn IMU to determine the frailty status of individuals.

For validating the performance of different sensor positions, the open-source algorithm proposed by Adamowicz et al. [Ada20a] was used for automated STS detection. While the algorithm has been validated for detecting STS transitions and for activities of daily living, such as drinking water or putting on a coat, its validation specifically for movements similar to STS transitions has not been conducted. As the algorithm was developed for an IMU placed on the lower back, an extension of the algorithm to other sensor positions and the assessment of its performance on different sensor systems is of interest. The chest position, for example, offers the potential to capture additional signals like an electrocardiogramm (ECG), aiding in the investigation of postural dizziness [Hap21]. The ear position offers potential advantages for long-term measurements due to its ubiquity, and the possibility of integrating the IMU into earbuds or hearing aids. Therefore, the algorithm was optimized and applied to data of IMUs placed on the lower back, chest, and ear position to compare their performance. Additionally, ML classifiers were trained and compared to further enhance the performance. These evaluations were conducted using an existing dataset recorded by a previous study containing data from young, healthy adults.

Furthermore, the optimized algorithm with the additional classifier for ear position was validated on a second existing dataset. This dataset consists of data collected from elderly, healthy adults using two different IMU systems. The performance of the two different sensor systems was compared.

Another objective of this thesis was to assess the frailty status using the detected STS transitions. Previous studies have revealed that frailty status can be assessed using automated STS detection by IMUs. However, most of the studies used standardized performance tests including STS transitions for frailty assessment. For that reason, this thesis aims to investigate the frailty status on STS transitions from both standardized tests and naturally performed transitions. The features extracted from the detected STS transitions, acquired by the STS detection algorithm, were used for assessing the frailty status in individuals. To achieve this, a study with elderly, frail participants was conducted. Frailty prediction was done by training and evaluating different classification pipelines and get the best performing one.

1.4 Outline

This thesis is structured as follows: Chapter 2 offers an overview of the fundamentals relevant to this thesis, including the evaluation of functional mobility, frailty, and its assessment, as well as an introduction to IMUs and traditional ML classifiers. The two previously existing datasets used in this thesis are described in Chapter 3, as well as the details of the data acquisition for the conducted study. Chapter 4 outlines the methods employed in this thesis, such as the STS detection algorithm, parameter optimization through grid-search, additional ML training, a comparison of two different IMU systems, the frailty prediction using a trained classifier, and the evaluation of all obtained results. The results and their discussion are presented in Chapter 5. Chapter 6 provides a summary of the discussion and presents the limitations of this work. Finally, Chapter 7 summarizes the findings, offers an outlook on further research questions, and concludes the thesis.

Chapter 2

Fundamentals

In this chapter, the fundamentals essential for this thesis are presented. First, the assessment of functional mobility and the evaluation of the health status of individuals is described, including an overview over functional performance tests, giving a description of frailty, and its assessment. Additionally, general information about IMUs is provided. Subsequently, a description is given for the four different ML classifiers used in this thesis.

2.1 Functional mobility and assessment

This section focuses on describing the medical fundamentals relevant to this thesis. First, an overview of the functional performance tests, which include STS transitions, is provided. Following that, the concept of frailty is introduced, encompassing the different approaches for defining frailty in practice, followed by an overview of frailty assessment methods commonly employed in clinical practice.

2.1.1 Functional performance tests

With people living longer and getting older, the likelihood of dependency on medical service, welfare, and other services increases [Mil14b]. The aging process is often accompanied by a decline in muscular mass and strength, which can result in several issues such as reduced mobility, increased risk of falls, frailty, and disability [Wal12]. Therefore, regular clinical assessment of muscle strength are crucial in evaluating therapy effectiveness and providing a reliable indicator of the severity of the disease process [Csu85]. Clinicians use a battery of tests based on activities of daily living to evaluate the physical function of elderly individuals [Mil14a]. While these tests provide qualitative information about the functional status of individuals [Atr20], they often lack in quantitative outcomes making it difficult to detect small differences and changes in movements.

One of the used movements within these tests is the STS transition. This transition, similar to the gait cycle, is a fundamental movement pattern [Nuz86]. Standing up and sitting down are daily tasks that require muscle power and balance [Atr20]. Difficulties in performing this task can be attributed to deficits in muscular strength or joint ranges of motion [Ale91]. Performance tests including STS transitions are used in clinical practice to gain insights into functional mobility and overall health status of individuals [Ada20a]. Analyzing the biomechanics of the performed STS transition during those tests allows for an investigation into the impact of natural aging and illness on immobility, as well as the identification of crucial factors that contribute to maintaining mobility [Ale91].

Standardized tests, such as the 30s CST and 5xSTS test, are commonly used to assess the ability to perform STS transitions and to evaluate the quality of the transitions [Atr20]. These tests provide valuable information on the functional status of older individuals [Cob20].

During the 5xSTS test, the time required to perform five consecutive STS transitions with arms crossed over the chest is measured, providing a quantitative measurement of lower body strength and functional capacity [Csu85; Gur94]. The 5xSTS is often performed as part of the Short Physical Performance Battery (SPPB), which also includes a walking test and several balance stand tests for assessing the healthy status [Gur94].

Another widely used test is the 30s CST, in which the number of STS transitions completed within 30 seconds, with arms folded over the chest, is recorded. This test is used to assess the functional status of older individuals, including the assessment of lower body strength and the detection of muscle weakness in participants. The evaluation of this test ranging from those who cannot complete a single stand (score of zero) to highly fit individuals who can perform 20 or more STS transitions during the 30 seconds [Jon99].

The TUG test is another functional performance test that combines standing up with a walking component. During this test, participants have to rise from a chair, walk three meters, and return to the chair while the time needed to complete the task is measured. The mobility level of participants is then categorized from unrestricted everyday mobility to distinct restrictions [Pod91].

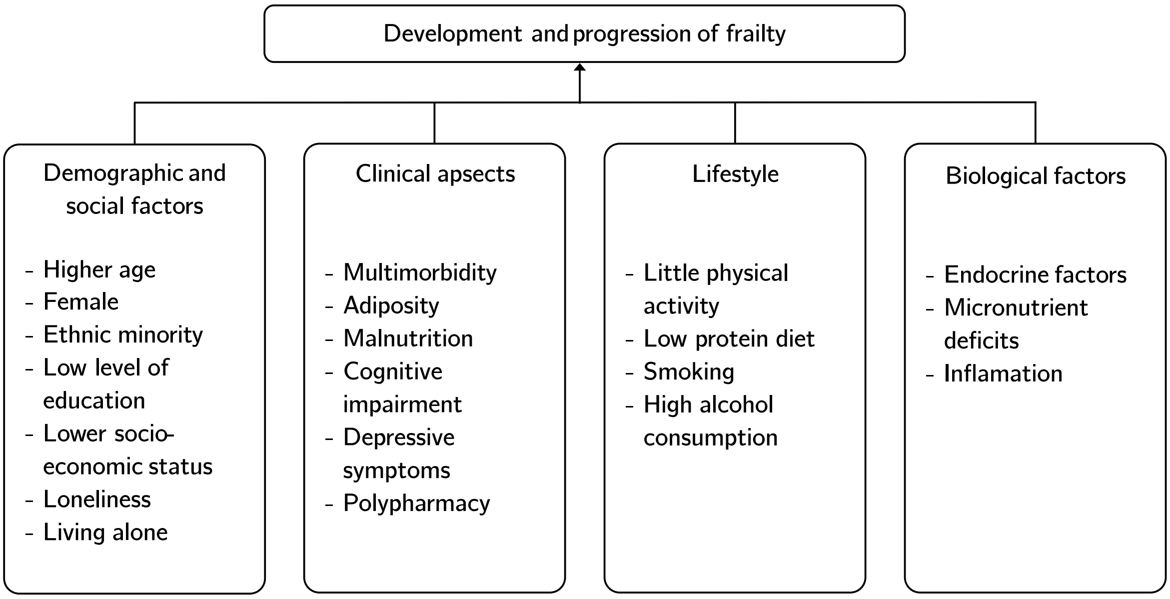


Figure 2.1: The different factors for the development and progression of frailty according to Benzinger et al. [Ben21]

2.1.2 Frailty

Frailty is a syndrome characterized by an increased vulnerability to both internal and external stressors, such as aging, disease, fitness, nutritional status, and other factors [Roc99]. It is associated with a decrease in the general health status [Bro10]. The development and progression of frailty involve various factors from different aspects, including demographic and social factors, clinical aspects, lifestyle, and biological factors, as shown in Fig. 2.1 [Ben21]. Elderly individuals affected by frailty face higher mortality rates and an increased risk of falls, hospitalization, and long-term care needs [Lac12].

In the field of frailty, there is no consensus regarding a uniform definition and assessment, leading to the emergence of different concepts [Ben21]. One prominent concept is the frailty phenotype proposed by Fried et al. [Fri01], also known as the physical frailty phenotype. The definition focuses primarily on assessing functional aspects, such as strength, mobility, and activity. With these aspects, the objective and subjective decline in performance capacity and weight loss were recorded. Fried et al. [Fri01] defined frailty as the presence of three or more of the following criteria: unintentional weight loss (shrinking), weakness (measured by grip strength), poor endurance and energy, slowness (walking speed), and low physical activity level. Based on these five criteria, individuals can be classified as robust (0 criteria are fulfilled), pre-frail (1-2 criteria are fulfilled), and frail (3 or more criteria are fulfilled).

Another concept, proposed by Rockwood and Mitnitski [Roc07], is the Deficit Accumulation Index (DAI) or Frailty Index. This approach considers the accumulation of 70 deficits, including functional, social, cognitive, and morbidity components. The Frailty Index is calculated as the proportion of potential deficits that are present in an individual, encompassing symptoms, signs, disabilities, diseases, and laboratory measurements. It was found that the coefficient of variation (CoV) of the frailty index decreases with age, indicating its relevance in capturing the frailty status of an individual.

2.1.3 Assessment of frailty

The frailty status of an individual can be used as a guide for clinical decision making and as a target for interventions [But16]. Various instruments are employed for the assessment of frailty, based on the two previously introduced concepts of frailty.

The frailty concept proposed by Fried et al. [Fri01] provides the basis for the widely used instrument physical frailty phenotype (PFD). This instrument involves the investigation of functional aspects of an individual, including the assessment of five criteria. With this tool, only the current state of an individual is assessed and it provides a high diagnostic effort with the test of walking speed and grip strength measurement additional to the questionnaire [Ben21]. Despite the increased effort, the PFD remains the gold standard for assessing frailty due to its validation across different geriatric settings [But16]. A more simplified instrument based on this frailty concept is the frail scale [Mor12]. It is easier to assess compared to the PFD as it evaluates frailty without clinical interventions. Four of the five questions are based on the Fried criteria. Additional diseases out of 11 prescribed were queried. The FRAIL scale can be assessed by medical personal or also by patients or relatives.

The Frailty Index, based on the concept of frailty defined by Rockwood and Mitnitski, is less commonly used in clinical practice. However, an instrument based on this index, known as the Clinical Frailty Scale (CFS), is used in various clinical settings [Ben21]. The CFS enables healthcare professionals to assess the degree of frailty of an individual [Roc05]. The assessment involves disease symptoms, functional decline, cognitive limitations, and life expectancy of an individual two weeks before the current acute illness based on an original seven point scale. Another concept based on the definition of frailty by Rockwood and Mitnitski is the Groningen Frailty Indicator, which includes 15 questions. Each item is dichotomized to calculate a score, where a higher score indicates the greater frailty of a subject with a maximum score of 15. The items cover physical, and cognitive impairments as well as psychosocial aspects. Additionally, there are other instruments available for assessing frailty, including the Gill Frailty Measure, Frailty/Vigor Assessment, Brief Frailty Instrument, Vulnerable Elders Survey, Frail Scale, and Winograd Screening Instrument [Ben21].

To summarize, frailty can be assessed using different concepts and instruments consisting of questionnaires and functional tests. The assessment provides insights into the progression of frailty, clinical significance, and potential interventions.

2.2 Inertial measurement unit

IMUs are small devices that incorporate multiple sensors to detect acceleration and angular rates of motion [Zha12]. There are two types of IMUs commonly used. The first type includes an accelerometer and a gyroscope, which measure acceleration and angular rate in three axes, respectively. The second type incorporates an accelerometer, a gyroscope, and a magnetometer, all measuring in three axes (Fig. 2.2). The magnetometer primarily enhances the data quality of gyroscope measurements [Ahm13]. IMUs with accelerometers and gyroscopes typically have three accelerometers and three gyroscopes mounted orthogonally to each other to enable measurements in three axes [Shk21].

IMUs find application in various fields, including quality control in the manufacturing industry, medical rehabilitation, robotics, inertial navigation systems, learning sports, and motion tracking in virtual reality systems [Ahm13].

Accelerometers were used in health care applications for measuring gait parameters, detecting falls, or recognizing activities [Ata11]. The acceleration can be obtained without the need for an external reference, as the acceleration of a rigid body is proportional to the force applied to it (Newton's second law) [Shk21]. The acceleration is typically measured in g-forces. Gyroscopes, on the other hand, measure the rotation [Shk21]. They are employed in various applications, such as automotive rotation detection, platform stabilization, gyrocompassing, and inertial navigation [Shk21]. The rotation is measured in degrees per second.

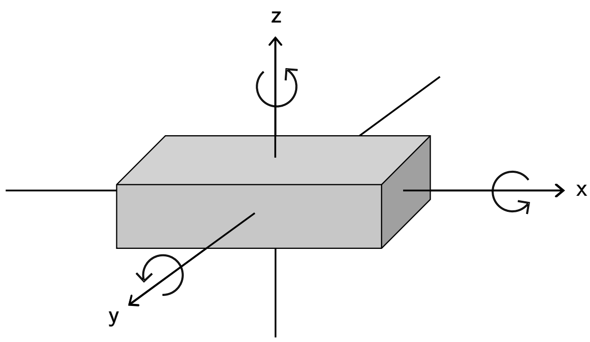


Figure 2.2: The axes in three directions for acceleration and gyroscope of an inertial measurement unit IMU

2.3 Machine learning classifiers

In the field of machine learning (ML), there are several approaches for training the models. Supervised learning involves providing the machine learning model with a training set consisting of input-output pairs. The goal is to train the model to generalize well to unseen data and avoid overfitting, where the model becomes too specialized for the training data. In contrast, unsupervised learning does not rely on labeled data. It aims to find patterns, structures, or relationships within the data without explicit guidance. The evaluation of unsupervised learning models is based on absolute error measures. These ML models generate output that can be grouped into numerical-continuous or categorical (discrete) values, which leads to different processes. For the first case, the process is called regression, while the second case is called classification [Bon17].

A pipeline, as shown in Fig. 2.3, is used for applying ML algorithms on data. Classifiers are built based on features extracted from preprocessed data. These features are then used as a training set to adapt the parameters of the classifier model. The classes of these features are known in advance. After training the model, it can be used for determining the classes of new input data, known as test set [Bis09]. The splitting of an existing dataset into test and training data can be done in various methods. One approach is using one part of the data for training and the remaining part for estimating performance. Another technique is performing cross-validation, where the training set is divided into mutually exclusive and equal-sized subsets. The classifier is then trained on the union of all other subsets for each iteration. Leave-one-out validation is a special case of cross-validation where each data point serves as a validation set [Mag07].

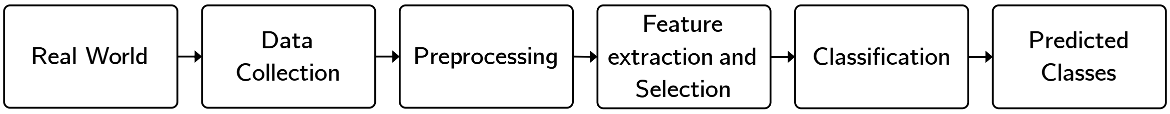


Figure 2.3: The steps of a machine learning (ML) pipeline according to Niemann [Nie83].

This section gives an overview of some existing ML classifiers. The introduced classifiers were used in this thesis for training the classifiers for optimizing the STS algorithm approach and for predicting frailty.

2.3.1 K-nearest neighbors

The k-NN is a classification algorithm that operates based on the principle of proximity. It assumes that dataset instances sharing similar properties tend to be closer to each other in the feature space and therefore tend to belong to the same class. An unclassified instance will then be classified by the algorithm by examining the class labels of its k-nearest neighbors. The algorithm identifies the class of this instance to the query instance based on the class label that most frequently occurs among these neighbors [Mag07].

2.3.2 Support vector machine

SVMs are a class of machine learning algorithms that use the concept of a margin, which is the space on either side of a hyperplane separating two data classes. By maximizing the margin, SVMs aim to achieve the largest possible distance between the hyperplane and the instances from each class, leading to a reduction in the expected generalization error. The hyperplane is defined by the closest samples from each class, known as support vector points [Mag07].

Linear SVMs are used for binary classification with linearly separable data, and the goal is to find the optimal separating hyperplane that maximizes the margin between two classes and therefore to reduce the probability of misclassifications. For non-linear problems, SVMs can leverage kernel functions to transform the feature vectors into a higher-dimensional space where the data become linearly separable again [Bon17].

2.3.3 Decision tree

Classification models that sort instances based on their feature values are called decision tree (DT)s. In a DT each node represents a feature and its corresponding instance to be classified. Each branch represents a possible value that the node can take. The classification problem starts at the root node and is sorted recursively based on their feature values [Mag07]. The construction of a DT involves training data consisting of objects described by attributes and class labels. The DT is built by finding a test that splits the data into subsets at each internal node [Mur98]. Binary DTs, specifically, operate through a sequential decision process. Starting from the root node, a feature is evaluated, and one of the two branches is selected based on the feature value of the instance. This process is repeated recursively as the tree is transversed until a final leaf node is reached, which represents the classification target [Bon17].

2.3.4 Random forest

A random forest (RF) is an ensemble of decision trees constructed using random samples from the data. Unlike decision trees that seek the best choice for splitting a node, RF employs a different strategy. It randomly selects a subset of features and tries to find the thresholds that separate the data best. This process is repeated for multiple trees in the forest, resulting in a set of weaker models, each with its prediction. Interpreting these results can be done using either a majority vote or by averaging the predictions from the individual trees [Bon17]. One advantage of RFs over DT is that they do not tend to overfit the data [Bre01].

Chapter 3

Datasets

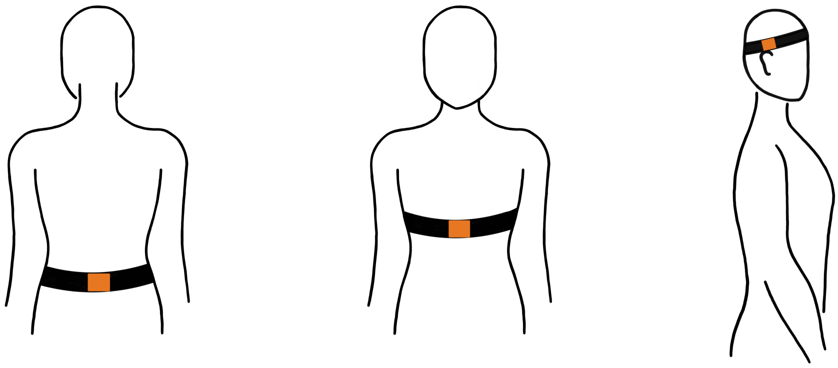
This chapter provides an overview of the data used for this thesis, which were analyzed to address the aims of this thesis. To accomplish this, three distinct datasets were employed. The first section of this chapter focuses on two datasets obtained from previous studies, providing a detailed description of each. In the second section, the third dataset is introduced, which was obtained through the conduction of a study. Tab. 3.1 provides an overview of the demographic data and conducted transitions for all three datasets.

3.1 Existing datasets

This section provides a detailed description of the two existing datasets used for analysis. The first dataset consisted of data from young, healthy adults who participated in a previous study, worn IMU on the lower back, chest, and ear position. The second existing dataset includes data from older, healthy adults, which wore IMUs on the ear.

3.1.1 Dataset 1: young, healthy adults

This dataset containing data from young, healthy adults will be referred to as *Dataset 1* in the thesis. The study involved the participation of 15 healthy, young individuals (8 males and 7 females; aged 25.6 $\pm$ 3.4 years; BMI 23.0 $\pm$ 3.6 kg/m^2). Before participation, their physical condition was assessed by the Physcial Activity Readiness Questionnaire (PAR-Q) [Tho92]. The study was approved by the local ethics committee of Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany (Re-No. 106_13B).



(a) Lower back (b) Chest (c) Ear

Figure 3.1: Sensor positions for Dataset 1.

For recording IMU data, three NilsPod sensors (Portabiles GmbH, Erlangen, Germany) were used, which recorded triaxial acceleration (range: $\pm$ 16 g) and triaxial gyroscope data (range: $\pm$ 2000 °/s) at a sampling rate of 256 Hz. To ensure synchronization, a wireless synchronization protocol [Rot18] was used and calibration was performed using the method introduced by Ferraris et al. [Fer94]. The sensors were attached to the lower back (L5 position), the chest (middle of the sternum), and over the right ear as shown in Fig. 3.1.

The study protocol involved performing five consecutive STS transitions on five different chairs and sofas, with a rest period of 20 seconds between each transition. Additionally, the participants performed different tasks of daily living, such as walking, climbing stairs, loading the dishwasher, getting items from the fridge, tying shoelaces, picking up objects, and lying down on a sofa. These tasks were selected to simulate movements in daily-life activities, which are similar to standing up from a chair. This selection aimed to ensure that the algorithm could distinguish between these tasks and actual STS transitions. The daily tasks were performed without receiving any further instructions on how to perform them.

During the data acquisition process, the different daily tasks and STS events were labeled using the smartphone application *aTimeLogger*[1]. This application allowed the configuration of different tasks and recorded the start and end times, as well as the duration, to serve as references. In total, each participant performed 54 STS transitions, which included the 5x5 transitions from different chairs and two daily task-related transitions, all repeated twice.

[1]https://play.google.com/store/apps/details?id=com.aloggers.atimeloggerapp

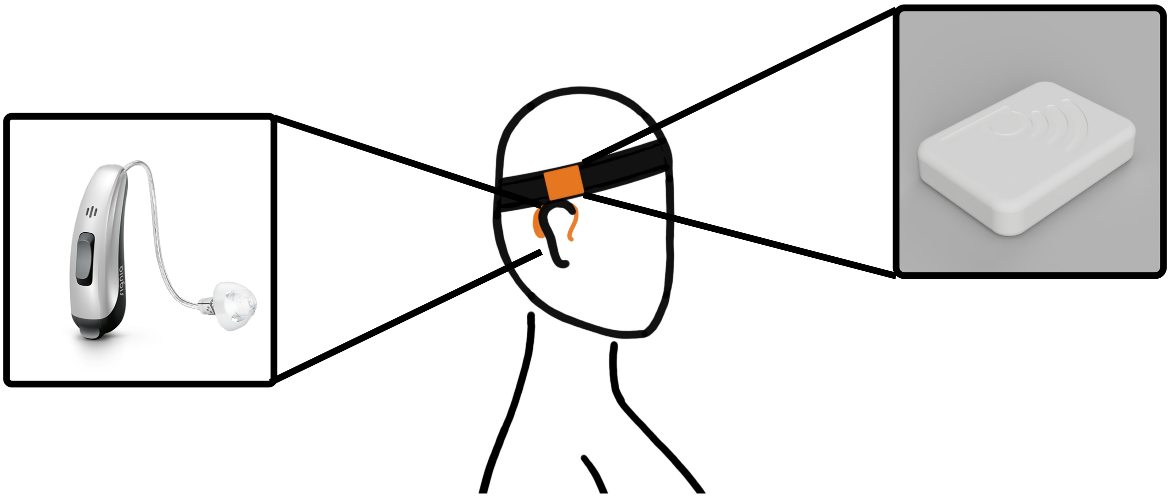


Figure 3.2: Position of the NilsPod sensor on the headband and of the Signia hearing aid with integrated IMU. These sensors were attached to both sides of the head.

3.1.2 Dataset 2: older, healthy adults

The second existing dataset, referred to as *Dataset 2*, comprises data from 21 healthy, elderly adults (10 males and 11 females; aged 70.9 ± 5.7 years; BMI 26 ± 3.1 kg/m^2). All participants were healthy with no limiting health conditions and provided written informed consent. The study received approval from the local ethics committee at FAU, Germany (Re-No. 106_13B).

For IMU data recording, two wearable NilsPod sensors (Portabiles GmbH, Erlangen, Germany) were used. These sensors recorded triaxial acceleration (range: ± 16 g) and triaxial gyroscope data (range: ± 2000 °/s) at a sampling rate of 204.8 Hz. Additionally, two Signia hearing aids with integrated IMU sensors (Sivantos GmbH, Erlangen, Germany) were used, which recorded triaxial acceleration (range: ± 2 g) and triaxial gyroscope data (range: ± 1000 °/s) at a sampling rate of 200 Hz. Both sensor systems were synchronized and calibration was done using the method introduced by Ferraris et al. [Fer94]. The IMUs integrated into the Signia hearing aids were placed in the right and left ear of the participants, while the two NilsPod sensors were attached over the right and left ear using a headband (Fig. 3.2). This setup resulted in four sensors for each participant, two on each ear.

To synchronize the two different sensor systems, participants performed synchronization movements at the beginning and the end of the measurement, involving nodding the head three times. During the measurement, participants performed the TUG test and the SPPB. These tests, which included walking tests, different balance tests, and a 5xSTS test, were described in Chapter 2.1.1. A video was recorded during the measurement to enable subsequent data labeling. For the labeling process, the Python package *MaDGUI* [Oll22] was used.

In total, each participant performed seven to 15 STS transitions during the measurement. These STS transitions included one or two transitions during the TUG test, depending on whether it was performed once or twice, and five transitions during the 5xSTS of the SPPB. Two participants performed the 5xSTS twice, resulting in 5 additional transitions. The number of additional transitions depends on how many times the participants sat down in between the tasks. Overall, 176 transitions were performed within Dataset 2.

3.2 Data acquisition: Dataset 3

In this section, the data acquisition for Dataset 3 is described, which was collected within a study conducted as part of this thesis. The section provides a description of the sensor setup, followed by an explanation of the study procedure. More detailed information about the comprehensive study design, protocol, and setup pictures can be found in Appendix A. Additionally, this section provides information about the individuals who participated in the conducted study.

3.2.1 Sensor setup

The objective of this study was to evaluate the frailty status of participants by analyzing data obtained from STS transitions using an ear-worn IMU. To collect this data, a Signia hearing aid (Sivantos GmbH, Erlangen, Germany) with an integrated IMU was used. The IMU sensor recorded triaxial acceleration (range: $\pm$ 2 g) and triaxial gyroscope data (range: $\pm$ 1000 °/s) at a sampling rate of 50 Hz. The sensor was attached to the right ear of each participant and was calibrated using the method introduced by Ferraris et al. [Fer94].

3.2.2 Study procedure

The study was divided into two parts, which were conducted on separate days with a two-day rest period in between. The study protocol was approved by the local ethics committee at FAU, Germany (Re-No. 23-66-S). The study protocol is shown in Fig. 3.3, the detailed study procedure including protocol and pictures can be found in Appendix A.

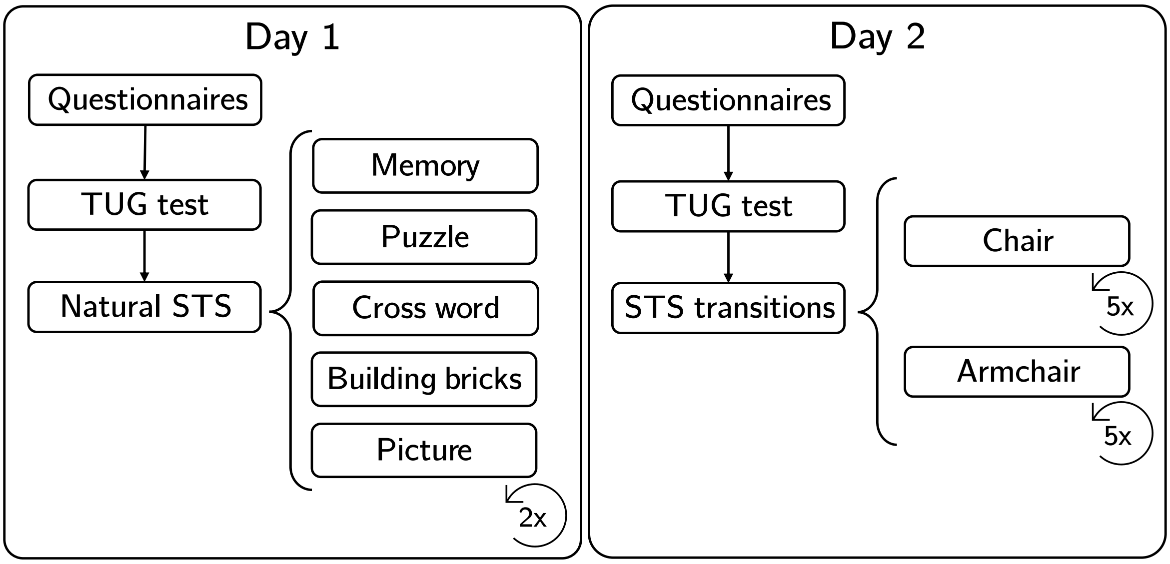


Figure 3.3: Protocol of the conducted study with older, frail participants. The timed-up-and-go (TUG) test was performed on both days, as well as different sit-to-stand (STS) transitions.

On the first day, participants received a brief introduction about the overall content of the study. After the introduction, several questionnaires were filled out in collaboration with the participants to assess their frailty status and current mood: Profile of Mood States [Dal02], Sarcopedia (SARC-F) [Mal13], Fried Frailty Phenotype (FFP) [Bra18], and Patient Health Questionnaire (PHQ-8) [Grä04].

Following the questionnaires, participants were provided with the hearing aid, and the data recording was started in the smartphone application of the hearing aids. The participants performed then the TUG test twice. Subsequently, the natural STS transitions were assessed by focusing on performing small tasks at five different stations. These tasks were solving a memory, completing a puzzle, building something with building blocks, filling out a crossword, and describing a picture. Each station was repeated twice, resulting in participants standing up 10 times during this study part. If participants were physically capable, the SPPB was conducted at the end. In total, the participants were standing up 7 to 16 times on the first study day, depending on the number of tests and stations performed.

After a two-day rest period, the second part of the study took place. The questionnaire aktuelle Stimmungsskala (ASTS) was repeated, and subsequently, the hearing aids were attached and data recording was started. The participants then performed the TUG test, twice. The main part involved participants performing five consecutive STS transitions from two different chairs with armrests. To provide rest periods between the transitions, participants were instructed to collect puzzle pieces after each STS transition and solve it during sitting.

In total, the participants performed 14 to 18 STS transitions during the second part of the study. These transitions included 10 transitions from the chairs, two transitions from TUG, and additional transitions in between the study parts. During both study days labels indicating the transitions were directly assigned to the IMU data via the smartphone application of the hearing aids. These labels served as references during the analysis.

3.2.3 Participants

In order to participate in the study, participants needed to meet certain criteria. They were required to be 60 years old or older and capable of standing up without assistance. Exclusion criteria included the inability to perform 15 STS transitions and having a care degree of five, which indicates the highest level of care required by an individual. Participation in the study was voluntary and all participants provided written informed consent before the study was conducted. Data were collected from 4 participants (1 male and 3 female) who did not have limiting health conditions regarding the study criteria. However, the participants exhibited a certain level of frailty, either pre-frail or frail. The age range was 80 to 91 years, with an average age of 86.8 $\pm$ 5.0 years. Their height ranged from 160 cm to 166 cm, with an average of 162 $\pm$3 cm, and their weight ranged from 48 kg to 70 kg, with an average of 60 $\pm$ 10 kg.

Table 3.1: Overview of the three described datasets with characteristics and demographic data. The timed-up-and-go (TUG) test and the short physical performance battery (SPPB) were performed in two datasets.

		Dataset 1	Dataset 2	Dataset 3
Demographic data	Number	15 (8 m, 7 f)	21 (10 m, 11 f)	4 (1 m, 3 f)
	Age	25.6 ± 3.4	70.9 ± 5.7	86.8 ± 5.0
	Height (cm)	176 ± 0.1	167 ± 8	162 ± 3
	Weight (kg)	72 ± 16	72 ± 10	60 ± 10
Characteristics	Young	✓		
	Old		✓	✓
	Healthy	✓	✓	
	Frail			✓
Sit-to-Stands	TUG		✓	✓
	SPPB		✓	✓*
	natural	✓		✓
daily tasks		✓		

* SPPB was only performed if participants were physically capable

Chapter 4

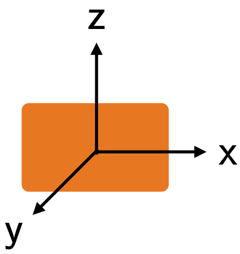
Methods

This chapter provides an overview of the methods used in this thesis. An overview of the preprocessing procedure is given, followed by a description of the STS detection algorithm by Adamowicz et al. used for detecting the transitions in all three datasets. In the subsequent section, the parameter optimization of the algorithm for the three sensor positions lower back, chest, and ear using grid-search is described. Additionally, the postprocessing step, involving feature extraction and training of classifiers to enhance the performance of the algorithm, is presented. This chapter further provides a description of the comparison between the two different IMU systems within the NilsPods and the Signia hearing aids, followed by a description of the training process for the frailty prediction classifiers. Finally, an overview of the evaluation methods used in this thesis is provided.

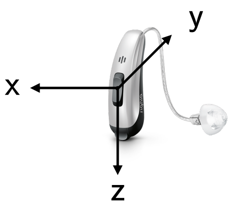
4.1 Preprocessing

To ensure proper alignment with the uniform coordinate system required by the used STS detection algorithm, preprocessing of the raw IMU data was done. As the sensors were placed on the three positions lower back, chest, and ear, and different sensor systems (NilsPod IMU and Signia IMU integrated into the hearing aid), every sensor had to be processed separately.

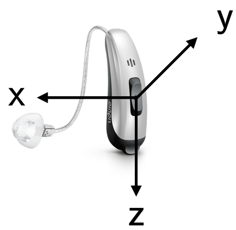
Each sensor was calibrated using the method introduced by Ferraris et al. [Fer94]. The first processing step was to apply the calibration on the raw data of each sensor. Subsequently, the coordinate system of each sensor was adjusted within its respective sensor frames through rotation. For the ear-worn sensors, an additional alignment to gravity was performed to orient



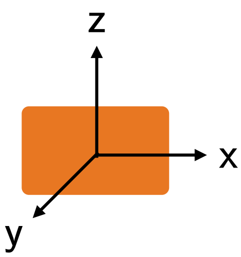
(a) NilsPod lower back



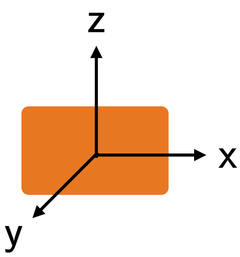
(b) Signia left



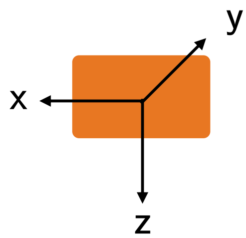
(c) Signia right



(d) NilsPod chest



(e) NilsPod left ear



(f) NilsPod right ear

Figure 4.1: Overview over the axes of all sensors after preprocessing.

them into the body frame. Since two different generations of hearing aids were used within Dataset 2 and Dataset 3, the axis alignment differed.

After this preprocessing step, the coordinate system of each sensor was aligned with the standardized coordinate system. The resulting coordinate system for each sensor is shown in Fig. 4.1. It forms a right-handed coordinate system, with the z-axis representing the vertical direction, the y-axis pointing to the front, and the x-axis to the left. Generally, the specific direction of the axis is not critical as the applied algorithm works independently of orientation.

4.2 STS detection algorithm

The performed STS transitions of the datasets were detected using the open-source algorithm by Adamowicz et al. [Ada20a]. This algorithm is available in the Python package *Sit2StandPy* [Ada20b]. It uses the vertical accleration data from the IMUand it provides the start and end time of the detected STS transitions, as well as the transition duration, maximum acceleration, minimum acceleration, vertical displacement, and spectral arc length (SPARC). Fig. 4.2 displays the filtered vertical acceleration of one STS transition of the three sensor positions lower back, chest, and ear. The by the algorithm detected STS interval is highlighted.

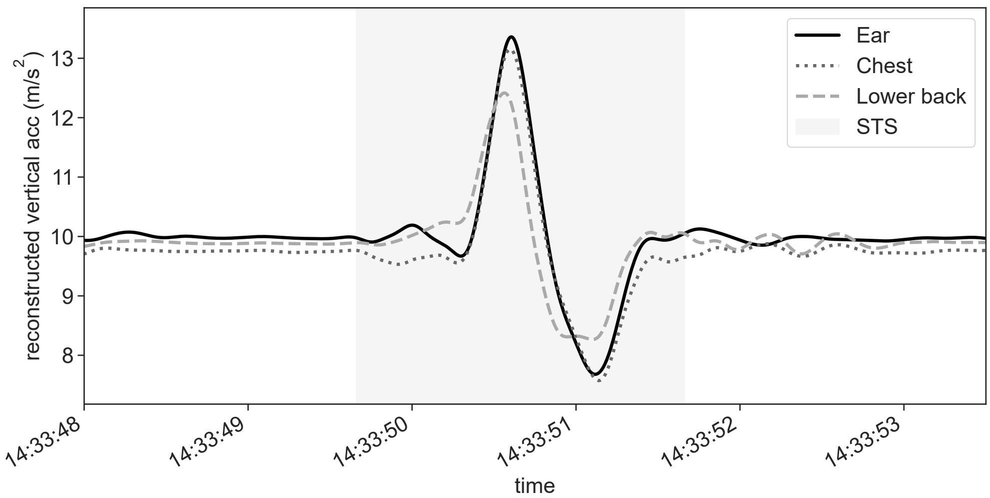


Figure 4.2: Reconstructed vertical acceleration of lower back position, chest position, and ear position as provided by the algorithm, detected sit-to-stand (STS) period is highlighted in grey.

The algorithm applies signal processing methods for detecting the transitions. Firstly, preprocessing of the raw acceleration data was performed by applying a fourth-order Butterworth filter with a cutoff frequency of 5 Hz. It is then smoothed by a rolling mean with a window of 0.25 s with uniform weighting. Next, a CWT is applied to the filtered acceleration data to identify potential intervals for STS transitions. This is done by summing the coefficients in the 0–0.5 Hz band at each time point for getting the CWT power. Within this power, the peaks are identified as possible time points for a transition.

In the second step of the algorithm, these potential STS transitions are validated. The periods of stillness within the filtered acceleration signal are identified by analyzing the rolling means and standard deviations. A set of predefined conditions is then applied to confirm or reject the potential transitions. These conditions are predefined and consist of criteria such as vertical velocity surpassing a specific threshold, a period of stillness required before the STS transition, the duration of the STS transition, the vertical displacement during the STS transition, and the duration between two consecutive STS transfers. These conditions serve as hyperparameters of the algorithm and can be modified using the application programming interface (API) provided by the Python package *Sit2StandPy*.

Table 4.1: Overview over the parameter with the default values and the parameter ranges of the sit-to-stand (STS) detection algorithm for grid-search.

Parameter name	Default	Grid
Acceleration moving average	0.2	$[0.10, 0.12, 0.14, ..., 0.24]$
Acceleration moving std	0.1	$[0.05, 0.10, 0.15, ..., 0.30]$
Displacement factor	0.75	$[0.65, 0.66, 0.67, ..., 0.85]$
Duration factor	10	$[6, 7, ..., 11]$
Jerk moving average	2.5	$[1.6, 1.7, ..., 3.5]$
Jerk moving std	3	$[2.0, 2.5, 3.0, ..., 6.0]$
Moving window	0.3	$[0.2, 0.3, 0.4, ..., 1.0]$
Transition velocity	0.2	$[0.05, 0.10, 0.15, ..., 0.4]$
Long still	0.5	$[0.4, 0.5, 0.6, ..., 4.5]$

Adamowicz et al. [Ada20a] provided default values for these conditions within the open-source package. These values were optimized for the performance of the algorithm using their own dataset, where an IMU was worn on the lower back. This provided parameter set will further be referred to as the *default* version of the algorithm.

4.3 Algorithm optimization for different sensor positions

In this thesis, the described algorithm was used and a further step was taken by optimizing the default parameter set specifically for the three sensor positions lower back, chest, and ear. To achieve this, experiments were conducted using a grid-search implemented using the Python package *tpcp* [Küd23]. This package enables the setup of a pipeline for computing the STS transitions for each sensor position and evaluating these results using the concept of *optimizable pipelines*.

To evaluate the detected STS transitions, the reference STS transitions were compared with the detected transitions provided by the algorithm. These references were used to determine true positive (TP) and false positive (FP) within the transitions, enabling the computation of metrics and validation of the detected STS transitions. A true detected transition was considered to be within a matching tolerance of two seconds around the reference time.

The grid-search for hyperparamter optimization was done on a high-performance cluster (HPC) of FAU. A list of possible values was defined for each adjustable parameter of the algorithm. The optimization process was performed using the F1-score as a metric. The range of each parameter used in the grid-search is provided in Tab. 4.1. Based on the grid-search results, optimized parameter sets were identified for each sensor position, improving the performance of the algorithm. Three different parameter sets were obtained: one for the lower back position, one for the chest position, and one for the ear position. These parameter sets will furhter be referred to as *optimized* version in this thesis.

4.4 Machine learning-based postprocessing of STS events

After optimizing the hyperparameters of the STS detection algorithm, it was evaluated whether the performance of can further be improved by a ML-based post-processing approach. The aim of this approach was to reduce the number of falsely detected STS events (FP). Thus, time-series features were estimated for each detected STS transition using the raw tri-axial acceleration and gyroscope data recorded between the start and the end points of the transitions as determined by the algorithm by Adamowicz et al. [Ada20a]. These features were extracted using the Python package *tsfresh* [Chr18] and served as input for training different classifiers to correct possible misdetections made by the STS detection algorithm. The ground truth labels, indicating true or falsely detected transitions, were determined by the references of the recorded datasets, respectively.

To identify the best pipeline in terms of classification performance, different combinations of preprocessing and classification algorithms were compared. Two preprocessing approaches were employed during this ML-step: scaling all features to a range of [0, 1] (*Min-Max Scaler*), or scaling the features to have a distribution with a mean value of zero and a unit variance (*Standard Scaler*). Four different classifiers were trained and investigated: k-NN, DT, support vector machine with linear basis function (SVM-lin) or support vector machine with radial basis function (SVM-rbf), and RF.

All the different combinations of pipelines were evaluated using a five-fold nested cross-validation (CV) with non-overlapping groups as employed in previous work for the evaluation of classification pipelines (e.g. [Ull23]). The evaluation process involved two loops: the inner loop performed hyperparameter optimization, while the outer loop conducted the training and evaluation of the classifiers. The F1-score was used as a metric for the optimization process. An overview of the parameter search space for each classifier can be found in Tab. 4.2. For the RF classifier, a randomized search with 1000 iterations was used, while a grid-search

Table 4.2: Overview over the selected parameter search space for the classifiers for the postprocessing of sit-to-stand (STS) events. For the random forest a random search with 1000 experiments was conducted, for the others a grid-search was used.

Classifier	Parameter	Value range
k-NN	neighbors weights	$[7, 10, 15]$ uniform, distance
Decision Tree	criterion max depth	$[gini, entropy, log_loss]$ $[8, 16]$
Random Forest	estimators criterion depth features min samples split max samples leaf	$[100, 200, 300...1000]$ $[entropy, log_loss, gini]$ $[8, 16]$ $[0.33, sqrt, log2]$ $[2, 5, 10]$ $[1, 2, 4]$
SVC linear kernel	C	$[-5, -4, -3, ..., 5]$
SVC radial basis function	C gamma	$[-5, -4, -3, ..., 5]$ $[scale, auto]$

was performed for the other classifiers. The Python packages *BioPsyKit* [Ric21] and *scikit-learn* [Ped11] were employed for training and evaluation of the ML algorithms. For the final classification, the classifier with the highest average F1-score across the outer CV folds was chosen independently for each sensor position. These classifiers were then used to predict the classes of the detected STS transitions.

4.5 Comparison of different sensor systems

In order to validate in the two-step approach described in the previous sections, which involves parameter optimization and the application of a classifier, Dataset 2 was used. The performance of the STS detection algorithm using this approach was evaluated on two different IMU systems: the NilsPod sensors and the IMU integrated into the Signia hearing aids.

Before applying the algorithm to the data of the two sensor systems, preprocessing steps were performed, which included synchronizing the two Signia sensors with the both NilsPod sensors to enable comparability. This was achieved by identifying the synchronization movement performed at the beginning and end of the data recording within both sensor data and the recorded video.

The synchronization point was defined as the moment when the chin reached the deepest position during the first trial of nodding. This point was identified in the gyroscope data of both sensors as a zero crossing from negative to positive on the x-axis. This axis corresponded to movements to the side. In the video data, the frames were obtained by identifying the first deepest point of nodding in both synchronization movements. Using these synchronization indices, the data from the NilsPod sensors were labeled using the *MadGUI* [Oll22] interface. Labels were set to identify the indices of the performed tests and STS transitions. These labels served as the ground truth for validating the detected STS transitions and analyzing the performance of the algorithm.

Following the labeling process, cross-correlation was performed on the data from both sensor systems. First, the lags between both signals were computed using the norm of the gyroscope data. Then, the correlation was calculated. With the correlation signals, the labels set in the NilsPod data can be used for the Signia data. The labels were used as a reference for validating the detected transitions. Due to the different sampling rates (NilsPod 204.8 Hz, Signia 200 Hz) and the resulting data drift, the matching tolerance was increased for the Signia data to four seconds around the reference time.

After completing these preprocessing steps for the signals, the algorithm by Adamowicz et al. with the optimized parameter set for the ear position and the additional classifier for this position was applied. The performance was compared between the left and right positions within each sensor system, as well as the performance between them.

4.6 Frailty prediction from STS features

To predict the frailty status, feature-based ML classifiers were trained using only the TP of the detected STS transitions which were identified based on the references events of the dataset, respectively. Feature estimation was performed using the Python package *tsfresh* [Chr18]. The features were trained using the raw acceleration and gyroscope data from start to end of the TP of the detected STS transitions, similar to the training of the classifiers for optimizing the algorithm. The extracted features provided by *tsfresh* were extended with the transition duration and feature selection was done. These selected features were used as input for training the classifiers that predict the frailty status.

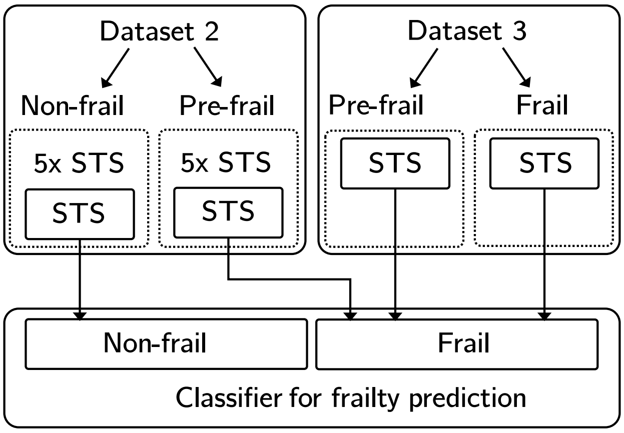


Figure 4.3: Overview over the used sit-to-stand (STS) transitions from Dataset 2 and Dataset 3 for training the classifier for frailty prediction and the categorization of the used transitions.

To obtain the required input data, frail and non-frail participants were needed. Since the participants within Dataset 3 were categorized into frail and pre-frail, additionally participants of Dataset 2 were used for getting this group. The frailty status was assessed based on the TUG time, resulting in non-frail, pre-frail and frail participants. To increase statistical power, the frail and pre-frail groups were combined into one for classification. For the participants included in Dataset 2, only the data from the right Signia sensor were used. The transitions performed during the 5xSTS within the SPPB were excluded from the prediction process, as they were fast, consecutive, which distinguished them from the other transitions. Additionally, participants in Dataset 3 did not perform the 5xSTS test. The used STS transitions from both datasets and the categorization of these are shown in Fig. 4.3. Consequently, this led to 12 participants within the frail group and 13 participants (8 from Dataset 2 and 4 from Dataset 3) within the non-frail group, based on the TUG time. For STS detection, the optimized parameter set for the ear position was combined with the trained classifier for this position.

Similar to Chapter 4.4, different preprocessing approaches and classifiers were combined to identify the best pipeline for frailty prediction. The min-max scaler and the standard scaler were used for preprocessing. The k-NN, the DT, the SVM-lin, the SVM-rbf, and the RF were trained and tested as classifiers. All pipeline combinations were evaluated using a stratified four-fold cross-validation with non-overlapping groups. The inner loop was used for hyperparameter optimization, while in the outer loop training and evaluation of the classifiers was done. The selection of the best-performing classifier was based on the F1-score. A randomized search with 1000 iterations was used for the RF classifier, and a grid-search was

Table 4.3: Overview over the selected parameter search space for the classifiers for frailty prediction. For the random forest a random search with 1000 experiments was conducted, for the others a grid-search was used.

Classifier	Parameter	Value range
k-NN	neighbors	$[2, 4, 6, ...20]$
	weights	uniform, distance
Decision Tree	criterion	$[gini, entropy, logloss]$
	max depth	$[2, 4, 8, ...20]$
Random Forest	estimators	$[100, 200, 300...1000]$
	criterion	$[entropy, logloss, gini]$
	depth	$[4, 8, 16]$
	features	$[0.33, sqrt, log2]$
	min samples split	$[2, 4, 6, 8, 10]$
	max samples leaf	$[1, 2, 4]$
SVC linear kernel	C	$[-5, -4, -3, ..., 5]$
SVC radial basis function	C	$[-5, -4, -3, ..., 5]$
	gamma	$[scale, auto]$

employed for the others. An overview of the parameter search space is provided in Tab. 4.3. The classifier that achieved the highest average F1-score across the outer CV folds was chosen for predicting the frailty status.

4.7 Evaluation

For the evaluation and validation of all used methods, several metrics were used, such as F1-score, precision, and recall. For all metrics, the detected STS transitions by the algorithm were validated with ground truth obtained from the reference labels in the different datasets.

Precision (also known as positive predictive value) serves as metric to evaluate the ability of the algorithm to detect the STS transitions. It was calculated as the ratio of TP of the detected STS to the sum of TP and FP, which corresponds to all detected STS events:

$$precision = \frac{TP}{TP + FN} \tag{4.1}$$

Recall (also known as sensitivity or true positive rate) is a metric that quantifies the ability to identify all performed STS transitions correctly. It is defined as the ratio of TP to the sum of TP and false negative (FN), which represents the total number of all performed STS events:

$$recall = \frac{TP}{TP + FN} \tag{4.2}$$

The F1-score, which is the harmonic mean of precision and recall, provides a balanced measure of the performance of the algorithm:

$$\textit{F1-score} = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{4.3}$$

Additionally, for evaluating the performance of the classifier for frailty prediction, accuracy was employed as metric. It is defined as the ratio of correctly predicted transitions (TPs and true negatives (TNs)) to all predictions:

$$\textit{Accuracy} = \frac{\#correct\ predictions}{\#total\ predictions} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.4}$$

Within Dataset 1, data from one participant were excluded from the data analysis due to failed sensor synchronization. Additionally, two participants had missing data from the ear-worn sensor. This resulted in data from 14 participants for the lower back and chest and in 756 performed STS transitions within these two sensor positions. For the ear-worn sensor, 12 participants remained, resulting in 648 performed STS transitions.

The comparison of the performance of the algorithm with the default and optimized parameter sets within each sensor position was performed using all available data, respectively. However, when comparing the three sensor positions lower back, chest, and ear, only the data from participants with data available from all sensors were used, resulting in 12 participants. The training and evaluation of the ML-based postprocessing step was also done using data from these 12 participants.

Dataset 2 was used for evaluating the performance of the algorithm on healthy, old participants. For this, data from all 21 participants were used, resulting in the analysis of 176 transitions. Additionally, the two different IMU systems were compared within the data of this dataset and the performance of the algorithm on other sensor systems was evaluated. The optimized parameter set and the best classifier for the ear position was used for this evaluation and analysis, as the sensors were placed there.

For evaluating the performance of the algorithm on older, frail participants, Dataset 3 was used. Data from all four participants from both study parts were included, resulting in 49 STS transitions for day one, 66 performed STS transitions for day two and in total 115 transitions.

The acquired Dataset 3, combined with the participants from Dataset 2, was used for the evaluation of the frailty status using the detected STS transitions. For assessing the frailty status the frail and pre-frail participants of both datasets were combined in one group. For the non-frail group transitions without the 5xSTS test of the non-frail participants from the Dataset 2 were used resulting in 13 participants. The TUG performance served as a reference for the frailty status in both groups.

Chapter 5

Results and discussion

In the following chapter, a comprehensive analysis and discussion of all obtained results is presented. Firstly, the results of the hyperparameter optimization are analyzed and discussed. This is followed by an examination and discussion of the trained ML approaches. Subsequently, the validation of the two-step approach on Dataset 2 is presented, along with a comparison of different IMU systems. Finally, the results of the trained classifier for frailty prediction are analyzed and discussed.

5.1 Algorithm optimization for different sensor positions

5.1.1 Results

The performance of the algorithm on the three IMU positions lower back, chest, and ear was evaluated using Dataset 1, which consists of young, healthy adults. First, the performance of the default and optimized parameter sets within each sensor position was compared, the results are shown in Tab. 5.1. When applying the STS detection algorithm to the lower back position using the default parameter sets, an F1-score of 97.3 % was achieved. However, after optimizing the hyperparameters on this position, the performance declined to 96.5 %. In contrast, using the optimized parameter sets on the algorithm for chest and ear position, both improved in performance. The F1-score for the chest sensor increased from 90.4 % to 91.7 %, while for the ear sensor it improved from 90.5 % to 95.1 %. Both increased F1-scores were obtained from improved recall in the optimized parameter set compared to the default parameter set.

Table 5.1: Performance metrics of the default and optimized parameter sets, including the number of true positives (TP), false positives (FP), and false negatives (FN). The highest F1-scores for each sensor position are highlighted in **bold**.

	Position	TP	FP	FN	Precision	Recall	F1-score
Default	Lower back	628	15	20	0.977	0.969	**0.973**
	Chest	549	17	99	0.969	0.847	0.904
	Ear	551	18	97	0.968	0.850	0.905
Optimized	Lower back	621	22	27	0.966	0.958	0.965
	Chest	568	22	80	0.963	0.877	**0.917**
	Ear	635	52	13	0.924	0.979	**0.951**

The F1-scores of the default and optimized parameter sets for the three different sensor positions are visualized as boxplots in Fig. 5.1. The plots include available data of Dataset 1 from all participants for each sensor, 14 participants for lower back and chest, and 12 for ear position. Therefore, comparison from the plot is only reliable within each sensor position. The boxplot for the lower back sensor with the optimized parameter set showed a higher variance compared to the one with the default parameter set. Additionally, an outlier is observed in the optimized results. Conversely, for chest and ear positions the boxplots for the optimized parameter sets decreased in variance compared to the default parameter set. The outliers of these positions are observed to be closer to the plots comparing the optimized parameter set with the default one for both sensors, respectively.

Second, a comparison of the computed metrics between the different sensor positions was done. Within this comparison, the lower back sensor maintained the overall highest F1-score with 97.3 % with the default parameter set. Regarding the performance metrics of the optimized parameter sets, the lower back sensor yielded also the highest F1-score with 96.5 %. The overall lowest performance with an F1-score of 90.4 % was recorded by the chest sensor within the default parameter set. In terms of the other metrics, the ear-worn sensor achieved the best overall recall with the optimized parameter set (97.9 %) and also the lowest overall precision of 92.4 %. After optimization of the parameter sets, the number of FP remained high across all sensor positions. Notably, the wrong detected STS transitions occurred predominantly during daily tasks, as shown in Tab. 5.2.

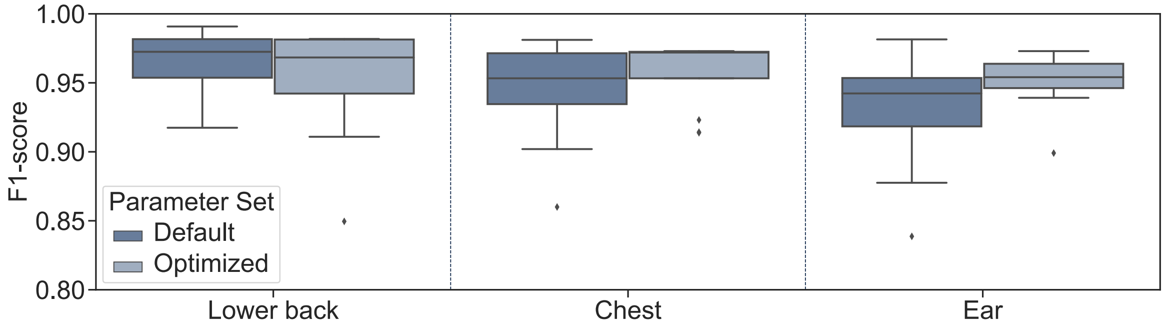


Figure 5.1: The boxplots were generated for default and optimized parameter sets, representing individual F1-scores for each activity of each participant for the three sensor positions. Outliers were omitted for visual clarity. The plots of lower back and chest position involved 14 participants, while the plot for the ear position included 12 participants.

5.1.2 Discussion

The first aim of this thesis was to optimize and validate an existing STS detection algorithm for the three sensor positions lower back, chest, and ear. A focus was on validating the performance of the algorithm on activities of daily living that involve movements similar to STS transitions. With implementing a grid-search, an optimized parameter set was found for each sensor position optimizing the performance of the algorithm for chest and ear positions. Additionally, the evaluation indicated a lack of robust STS detection in real-world activities.

The algorithm by Adamowicz et al. [Ada20a] was originally developed for the lower back position. When applying this algorithm with the provided parameter set to other sensor positions, a decrease in performance was observed compared to the original lower back position. To address this issue, a grid-search was employed to optimize the parameter set for all three sensor positions. This optimization led to improved performance of the algorithm for chest and ear position, resulting in more detected STS transitions. However, for lower back position the default parameter set outperformed the optimized parameter set.

A possible explanation for this could be that the algorithm by Adamowicz et al. was already optimized for this position and the provided parameter set was already tuned on F1-score for this position within their own dataset. This assumption is supported by the fact that the F1-score of the lower back position using the default parameter set achieved the best overall performance. In contrast to this, the F1-scores from the chest and ear position improved after optimizing the parameter sets. This improvement in F1-score was primarily

Table 5.2: Median and inter-quartile range (IQR) of the precision for the performed activity of daily living in comparison to the chairs for the optimized parameter set of each sensor position. The lower back and chest position involved 14 participants, whereas the ear position included 12 participants.

Position	Activity	Median	IQR
Lower back	Chairs	1.0	0.0
	Daily tasks	1.0	0.5
Chest	Chairs	1.0	0.0
	Daily tasks	0.75	0.5
Ear	Chairs	1.0	0.0
	Daily tasks	0.75	0.5

based on an enhanced recall score due to fewer FN, respectively. This indicated that more real STS transitions were detected after optimizing the parameter set for both positions.

A closer examination of the performance within the different performed activities revealed that the algorithm performed well for detecting the STS transitions of the different chairs and sofas within all three sensor positions, showing few FN and no FPs. However, during the simulated daily-life activities with STS-near movements, the algorithm showed a lack of accuracy. All falsely detected transitions occurred during these daily tasks. This resulted in a concern about the reliability of the STS detection algorithm within real-life situations and unsupervised measurements.

Adamowicz et al. [Ada20a] achieved a recall of 94.7 % and a precision of 99 % within their conducted lab study with data from young, healthy adults and the sensor placed on lower back. Additionally, within the lab-based study on participants with PD, the algorithm achieved a recall of 85.3 % and precision of 98.8 %. Comparatively, the best performing sensor within Dataset 1, which was the lower back sensor with the default parameter set, showed similar results with a recall of 96.9 % and precision of 97.7 %. However, there were differences in the protocols between the study from Adamowicz et al. and the study for Dataset 1. Both studies conducted by Adamowicz et al. involved participants performing the 5xSTS test with few additional natural STS transitions, while Dataset 1 contains participants performing five consecutive normal STS transitions.

Overall, the results showed that the algorithm by Adamowicz et al. [Ada20a] could be applied to different sensor positions. However, it is necessary to adjust the parameter sets individually for each position to achieve optimal performance. With this parameter adjustment, the algorithm showed good performance within all positions. Each sensor position can be used for specific monitoring tasks. Using a sensor worn at the chest provides the benefit of capturing ECG measurements enabling health analysis, such as postural dizziness [Hap21]. Ear-worn IMUs, on the other hand, are promising for real-life assessment due to the possibility to integrate them into commonly used devices such as hearing aids or earbuds.

5.2 Machine learning-based postprocessing of STS events

5.2.1 Results

Following the optimization of the hyperparameters for the STS detection algorithm for each sensor position, the next step involved training classifiers to further enhance the performance of the algorithm. The best-performing parameter sets, based on the F1-score, were selected for each sensor position to be used for the ML-based post-processing step of the detected STS events. Therefore, the default parameter set was used for the lower back position, while the optimized parameter sets were employed for the chest and ear positions, respectively. Tab. 5.3 provides an overview of the performance metrics for all three sensor positions, both before and after classification.

Among the four evaluated classifiers, the RF classifier showed the highest F1-score for the lower back sensor, increasing it from 97.3 % to 98.4 %. This improvement in the F1-score was primarily attributed to an increased precision from 97.7 % to 100 %. The recall score for the lower back remained unchanged. For the chest sensor, the SVM-rbf was the best performing classifier. A similar F1-score was achieved compared to the metrics before the classification step (91.8 % vs. 98.9 %). The improvement in F1-score comes from a slightly increased precision score by 0.1 %, while the recall score remained the same. Lastly, for the ear sensor, the DT classifier yielded the best performance. It achieved an F1-score of 98.9 %, which was 3.8 % higher than before applying the classification step. This improvement can be derived from an increase in precision from 92.4 % to 100 %, similar to the lower back position. Consequently, the ear position exhibited the highest performance in F1-score among all positions after the classification step.

Table 5.3: Performance metrics before and after the classification of the detected sit-to-stand (STS) transitions of the three sensor positions. For this comparison, data from 12 participants available in all positions were used.

Position		Precision	Recall	F1-score
Lower back	Before	0.977	0.969	0.973
	After	1.000	0.969	0.984
Chest	Before	0.963	0.877	0.917
	After	0.964	0.877	0.918
Ear	Before	0.924	0.979	0.951
	After	1.000	0.979	0.989

5.2.2 Discussion

After optimizing the algorithm on the three sensor positions lower back, chest, and ear position, it was found that the algorithm was not robust in detecting STS-like movements during activities of daily living, resulting in a high number of FP within all sensor positions. To address this issue, a post-processing step was incorporated into the algorithm using the best-performing parameter sets for each sensor position. A ML-based classifier was trained specifically for each sensor position to improve precision and reduce FPs.

The combination of the default parameter set and the classifier achieved 100 % precision for the lower back position. This indicated that all falsely detected STS transitions (FP) were eliminated successfully through classification. Similarly, the ear position achieved also a precision of 100 % with the optimized parameter set combined with the best classifier for this sensor position. This resulted in the ear position being the best performing sensor position overall in terms of all three metrics, precision, recall, and F1-score. The precision for chest position also improved slightly, but there remained FPs after applying the classifier on the algorithm with the optimized parameter set.

The achieved F1-score of over 98 % for lower back and ear position was higher compared to the approach of the multilayer perceptrons-based classifier by Hellmers et al. [Hel19]. They achieved an F1-score of 94.8 % using an IMU worn on the lower back with older participants performing the 5xSTS test. However, direct comparison is limited due to the different population characteristics and differently performed STS transitions. In a study with comparable activities to Dataset 1, Zhang et al. [Zha14] achieved a similar F1-score of 96 % using a SVM-based algorithm. However, the participants were older and a pendant IMU was employed for data recording.

Despite the improvements achieved by applying the ML classifier on the optimized parameter sets in this thesis, there were still FN for all three sensor positions, indicating some performed STS transitions were not detected by the two-step algorithm. It is important to note that the two-step algorithm cannot reduce the number of FN since transitions undetected by the algorithm in step one do not proceed to the classification step. However, this trade-off was considered acceptable as the primary goal was to enhance the precision and eliminate the FP in order to accurately assess the health status based on the detected STS transitions. This is important in unsupervised and long-term measurements where the reliability of algorithm results is crucial to avoid falsifying the health status with FPs.

The incorporation of the classification post-processing step into the algorithm and combining it with the best-performing parameter sets resulted in more robust outcomes, particularly for lower back and ear position. This makes the algorithm promising for real-world measurements using the lower back and ear positions.

5.3 Comparison of different sensor systems

5.3.1 Results

For validation of the developed algorithm with the additional classifier step on a different population than young, healthy adults, Dataset 2 was used containing older, healthy adults. Additionally, a comparison of two different sensor systems was performed, as Dataset 2 contains data from NilsPod IMU and Signia IMU (hearing aid integrated IMUs). The optimized parameter set, along with the best classifier of the ML process for the ear position, was used to analyze the data.

The results of all four sensors are displayed in Tab. 5.4. When comparing sensors of the same system on the left and right ear, the ear-worn NilsPod sensors exhibited similar performance. The differences in TP, FP, and FN resulted in differences of around 1 p.p. in all metrics. Specifically, the left sensor showed lower recall and the right sensor achieved lower precision and F1-score. Comparing the left and right Signia sensors resulted in the

Table 5.4: Performance metrics of all tested sensors including the NilsPod sensor on the left and right position and the Signia sensor on the left and right position. The number of detected true positive (TP), false positive (FP), and false negative (FN) are listed. The sum of transitions performed by all 21 participants is 176.

Sensor	Position	TP	FP	FN	Precision	Recall	F1-score
NilsPod	Left	148	29	28	0.836	0.841	0.838
	Right	149	32	27	0.823	0.847	0.835
Signia	Left	142	46	34	0.755	0.807	0.780
	Right	140	40	36	0.778	0.795	0.786

same differences. Recall was 1.2 p.p. lower for the right sensor, whereas precision was 2.3 p.p. lower and F1-score was 0.6 p.p. lower for the left sensor.

The comparison between the different sensor systems indicated a slightly higher performance for the NilsPod sensors within all metrics, including higher numbers of TPs, lower numbers of FNs, and FPs. Especially the number of FP was higher for both Signia sensors compared to the NilsPod sensors. The left NilsPod demonstrated the lowest number of FP, whereas the left Signia sensor showed the highest falsely detected transitions.

5.3.2 Discussion

Dataset 2 was used for validating the previously developed two-step approach of the STS detection algorithm for ear position. Results showed a lower performance compared to the results obtained from the ear position within Dataset 1. This difference in performance could be attributed to the fact that the algorithm was developed and optimized using data from young adults, whereas Dataset 2 contains data from older adults. The two age groups display variations in their standing up behavior. Due to typically having lower muscle strength, older adults tend to employ an increased trunk movement as a compensatory mechanism when standing up [Wal12].

Besides the validation, a comparison between two different IMU systems was done using the data from Dataset 2. When comparing the left and right sensors within each system, similar performance metrics were obtained. The differences observed were minimal, with only 1 p.p. differing in metrics for both the left and right NilsPod sensors. However, the comparison results for the Signia sensors showed higher differences. Overall, both systems demonstrated similar performance with a slightly higher F1-score of around 5 p.p. achieved by the NilsPod sensors. Similar findings were reported by Abdollah et al. [Abd21], where different IMU systems were used on chest and head positions during study conduction. Preliminary investigations indicated no differences between the sensor readouts when placed on the same body part.

The slight differences in performance within left and right sensors could be attributed to head tilting during the STS transitions. As the advisor stood to the right or left side of the subject during the examination, if the participant turned towards the examiner during the STS process, it could have led to variations in the signals captured by the left and right sensors, resulting in slightly different transition detection by the algorithm.

Overall, the performance of the Signia sensors was slightly lower compared to the Nilspod sensors, despite minimal differences in sensor positions on both ears. This discrepancy in performance was possibly due to the use of NilsPod data for setting the labels for the references. Although the Signia data were correlated with the Nilspod data to incorporate the references, the different sampling rates (NilsPod 204.8 Hz, Signia 200 Hz) may have caused a mismatch in the timing of the references within the Signia data. This was tried to be compensated by increasing the matching tolerance in the validation of the transitions. However, it cannot be ruled out that this discrepancy contributed to an increased number of FPs.

Examining all occurring FP more closely, no specific pattern or prominent movements were identified within the detected transitions classified as FP. Video recordings of the participants revealed that head movements particularly increased nodding during the instructions, were noticeable. This increased nodding may have contributed to the higher number of FPs observed across all sensors. It is worth noting that the data collection for Dataset 2 took place during the period of mandatory mask-wearing during the COVID-19 pandemic. In comparison to Dataset 1, which was recorded without masks and did not exhibit remarkable FPs in the ear sensor data, the presence of masks could potentially amplify gestures and head movements to compensate for communication limitations. A study has indicated that found an increase in hand gestures and movements during a speech when wearing masks [Kha22]. Additionally, another study suggests the use of body language as an assistance to verbal communication, including appropriate head movements such as tilting to the side or nodding [Mur21].

Table 5.5: Overview over the number of performed transitions on day 1 and 2 and in a total of the study, with true positives (TP), false positives FP, and false negatives FN, and the corresponding metrics.

	Transitions	TP	FP	FN	Precision	Recall	F1-score
Day 1	49	17	1	33	0.944	0.340	0.500
Day 2	66	10	8	56	0.556	0.152	0.239
Total	115	27	9	89	0.750	0.233	0.355

5.4 Frailty prediction from STS Features

5.4.1 Results

STS detection

First, the performance of the STS detection algorithm on Dataset 3 with old, frail participants was investigated using the algorithm with the optimized parameter set in combination with the best-performing classifier for ear position. The results were listed in Tab. 5.5. The algorithm detected successfully 27 transitions out of 115 performed transitions within the two study days, resulting in a low detection rate of 23.5 %. Overall, 36 transitions were detected by the algorithm, whereas 9 transitions were identified as FP.

On the first study day, 17 out of 49 performed transitions were detected by the algorithm as TP, and one additional false transition was detected. However, the performance of the algorithm within the second day was lower, with 10 TP and additional eight FP detected transitions out of 66 performed transitions. The overall F1-score was low with 35.5 %, primarily due to the low recall of 23.3 %.

Frailty prediction

For training the classifier to predict frailty, only the true detected transitions were used. This resulted in a dataset consisting of 83 transitions including 47 transitions from frail participants, including 20 transitions from Dataset 2 and 27 transitions from Dataset 3. Additionally, 36 transitions from non-frail participants, all from Dataset 2, were included as input for the

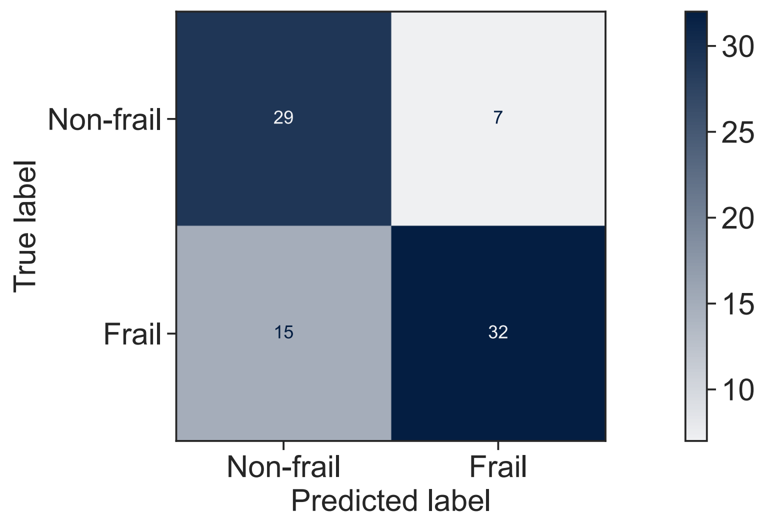


Figure 5.2: Confusion matrix yielded by the classifier for frailty prediction.

classifiers. Among the trained classifiers, the RF performed best, achieving a precision of 82.1 % in predicting frailty. Only seven non-frail transitions were falsely classified as frail transitions. Overall, 32 transitions out of the 47 were predicted as frail, resulting in 15 FN. Therefore, the recall was 68.1 %, and the F1-score was 74.4 %. The results are shown in the confusion matrix in Fig. 5.2, demonstrating an overall accuracy of 73.5 % in correctly classifying the transitions. Specifically, the accuracy of correctly identifying frail transitions was 68.1 %, while for non-frail transitions, 80.6 % were classified correctly.

5.4.2 Discussion

STS detection

Validating the performance of the algorithm on Dataset 3 with frail participants revealed a low recall rate of 23.5 %. This indicates that a high number of performed STS transitions were not detected by the algorithm, indicating limitations in accurately identifying all performed transitions within frail participants. However, the algorithm showed better performance on day one of the study compared to day two, particularly in terms of FPs, which contributed to a good precision rate of 94.4 %. Therefore, within the first day, one can rely more or less on the fact that a detected STS transition was a real one. Unfortunately, many transitions were missed throughout the recording, resulting in a low recall rate.

The better performance within data from day one could be attributed to the participants performing better transitions with rest periods in between, during which the small tasks were performed. Additionally, the focus was not solely on the transitions. On day two, the participants performed the STS transitions faster and consecutively, requiring greater effort. This could contribute to the differences in the performance of the algorithm compared to day one. Although attempts were made to incorporate rest periods on day two, they were apparently not sufficient.

Frailty prediction

The initial plan was to apply the classifier for frailty prediction on the performed STS transitions identified by the algorithm within data of Dataset 3, with the assessed Fried Frailty Index serving as the variable to be predicted. However, this approach was rejected due to the small number of participants in Dataset 3 and the absence of non-frail individuals in the study. Difficulties in recruiting participants from the retirement home resulted in the few participants, and establishing further collaborations within the given time frame was not possible. Consequently, Dataset 2 was additionally used to obtain the non-frail group for frailty prediction. Nevertheless, the Fried Frailty Index was not acquired for Dataset 2. Therefore, the duration time of the TUG test was employed to classify all participants from both datasets (Dataset 2 and Dataset 3) into frail, pre-frail, and non-frail categories. For training the classifiers for frailty prediction, the frail and pre-frail participants within Dataset 2 and 3 were combined into one group.

The trained classifier demonstrated good performance in predicting frailty based on the detected STS transitions. It achieved a good precision rate, indicating a low number of transitions falsely classified as frail. However, due to the small number of participants and therefore a low number of transitions used for training the classifier, results have to be treated with caution. The findings have to be seen as an initial approach to explore the differentiation between the frail and non-frail groups. Additionally, a visual representation of the four most important features in Fig. 5.3 further supports the findings of the classifier. The graphic indicates differences in the assessed features from the STS transitions between the two groups. Specifically, it reveals that the non-frail transitions tend to display a more uniform pattern, whereas the frail transitions are more scattered. This emphasizes the ability of the classifier to predict the frailty status based on automatically detected STS transitions.

The accuracy rate for correctly identifying the frailty status of a transition was 73.5 %. Similar results were achieved by the classifier trained by Greene et al. [Gre14], which achieved an accuracy of 72.88 %. In this study, a logistic regression model was trained for frailty prediction using STS transitions obtained by the TUG test.

Another study conducted by Millor et al. [Mil17] focused on frailty prediction using a classification tree model and kinematic parameters. Their results showed an accuracy for the non-frail group of 100 % and 70.1 % for the frail group. Participants were categorized into non-frail, pre-frail, and frail groups, and transitions of the 30s CST were used for classification.

However, direct comparison between the studies is limited due to the different tests used to obtain STS transitions for predicting frailty. Furthermore, Millor et al. [Mil17] compared the metrics specific to the different groups and did not provide an overall evaluation like in this thesis. In this thesis, the accuracy of correctly frail detected transition was 68.1 % and 80.6 % of the non-frail transitions were classified correctly. These results demonstrate similar performance within the frail group compared to Millor et al. [Mil17]. However, in this thesis, the pre-frail and frail participants were combined into one group.

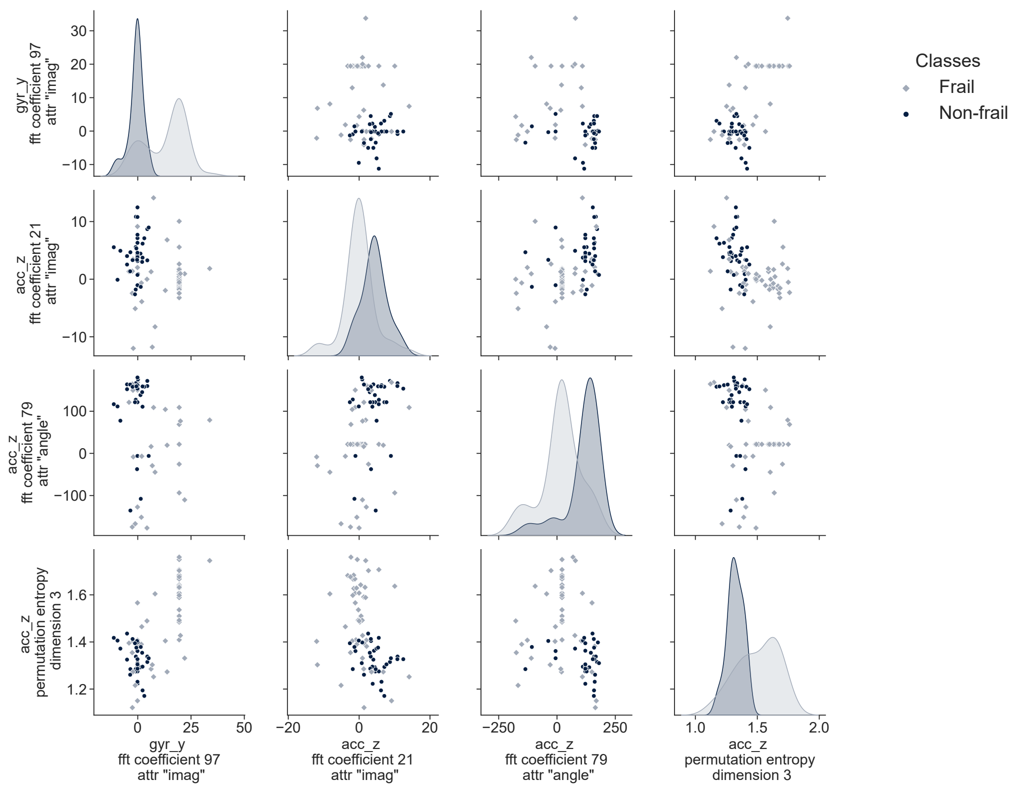


Figure 5.3: The four most important features according to the selected features of feature extraction of the detected sit-to-stand (STS) transitions for frailty prediction.

Chapter 6

Discussion and limitations

6.1 General discussion

The sit-to-stand (STS) detection algorithm developed by Adamowicz et al. [Ada20a] was used within this thesis for the automatic detection of transitions. In a previously conducted study, the algorithm with the default parameter set was applied to the sensor positions lower back, chest, and ear to evaluate the performance. However, the two additional sensor positions performed worse compared to the lower back. This was because the parameter set of the algorithm was originally optimized to the lower back position. Therefore, a grid-search was conducted in this thesis to optimize the parameter sets for the three sensor positions lower back, chest, and ear. The results showed better performance for chest and ear sensors using the optimized parameter sets on the algorithm. In contrast, the lower back sensor remained better performance with the default parameter set. This was because the default parameter set was developed for this position. Overall, the results showed that the algorithm by Adamowicz et al. can be applied to different sensor positions when properly adjusted.

Further analysis of the results from the best-performing parameter sets focused on the FP. It was found that these false detected transitions occurred especially in the performed daily tasks, which were selected because of STS-near movements. These results limit the use of this algorithm in unsupervised or real-world measurements. Therefore, a second ML-based postprocessing step was applied to the algorithm. This classification step improved the performance on the lower back and ear positions, eliminating all FPs, whereas the ear position performed best comparing all three sensor positions.

The ear position is beneficial for use in long-term measurements because they can be integrated into devices many people already wear during their daily-life activities such as headphones or hearing aids [Röd22]. Therefore, validation of this two-step approach was done for ear position using Dataset 2, which included a population of healthy older individuals. Additionally, a comparison of different IMU systems was performed. The NilsPod sensors performed better compared to the Signia sensors. This could be due to the references set up based on labels from the NilsPod data. The findings could be attributed to the development of the two-step approach on the NilsPod sensor worn on the right side over the ear.

The performance STS detection algorithm with the two-step approach of the ear-worn IMU was validated on data of older people, both healthy and frail. The algorithm performed worse for Dataset 2 and Dataset 3 compared to the results of the ear position in Dataset 1. This was due to a high number of FPs and FNs, emphasizing both missed true transitions and false detected transitions. These discrepancies in performance could be attributed to differences in motion patterns between the older adults and the younger ones. Older adults, who typically have lower muscle strength, tend to stand up with more dynamic trunk movements [van18]. This increased trunk use may be a compensatory mechanism due to the progress of muscular decline during aging [Wal12]. Consequently, the algorithm performed worse by applying it to data of older adults, as it was initially developed on young, healthy individuals.

These findings were also observed when visually analyzing the video recordings of Dataset 2. Especially during the 5xSTS, participants exhibited more trunk movement and momentum when standing up. This behavior could be due to the compensation of lower leg strength. Similar findings were also observed in Dataset 3. For the older frail population, the algorithm detected only few STS transitions, resulting in a high number of FNs. The performance of the algorithm was worse for Dataset 3 compared to Dataset 2.

The detected transitions from participants in Dataset 2 and Dataset 3 were used for training the classifier for predicting frailty, combining the pre-frail and frail to one group. By considering only the TP of the detected STS transitions, the classifier successfully differentiated between frail and non-frail transitions, achieving a good precision rate and indicating few misclassified transitions. This difference in the two groups was further confirmed by a graphical analysis that revealed differences in features between the two groups, especially in the distribution within both groups. Due to the limited dataset, these results indicate an initial approach for predicting frailty with the two-step approach developed by the algorithm.

6.2 Limitations

The used datasets for the investigations all have limitations as they were collected in laboratory environments. However, attempts were made to get more natural settings. In Dataset 1, daily-life activities were included within data recording to capture more natural movements. However, performing five consecutive STS transitions followed by standing was deemed unnatural. In real-live scenarios, participants would typically engage in activities like walking after standing up. Within Dataset 2, only standardized tests were performed by participants. The TUG test involved walking immediately after the transition, which aimed for a more natural transition. However, the transitions during 5xSTS test differed from a typical STS transition. Dataset 3 provided the most natural setting, with small tasks to be completed and the STS transition serving as a means to reach the tasks on day one.

The optimization of the parameter sets and the additional classification step were performed using Dataset 1. However, the lower performance of the algorithm on the other two datasets, which consisted of different populations compared to Dataset 1, indicates that the two-step approach may have been overfitted to the movement patterns of young, healthy adults. This limitation may restrict its applicability to other populations, such as older adults who exhibit different movement patterns during standing up, including more trunk movements.

Regarding the frailty prediction, the small number of participants available for training the classifier represents a limitation. Only 12 participants were classified as frail, with eight from Dataset 2 and four from Dataset 3. Additionally, Dataset 2 contributed 13 participants classified as non-frail. This limited sample size restricts the reliability of the results. Moreover, the frail group comprises participants from two different datasets, with participants of Dataset 3 performing a higher number of STS transitions, while the non-frail group consisted of more participants, who performed fewer STS transitions. Another limiting factor is the low number of performed transitions detected within Dataset 3. However, there were also few FP detected transitions, indicating the reliability of the detected transition. Therefore, the results of the classification and their interpretation should be treated with caution.

Overall, the frailty prediction can be seen as theoretical approach to investigate the feasibility of such a prediction. The results obtained by the classifier, along with the graphical representation showing the separation of frail and non-frail classes, provide initial insights into the potential of the approach. The next step would involve acquiring additional data from both frail and non-frail participants to further validate and evaluate the approach.

Chapter 7

Conclusion and outlook

This thesis addressed several objectives. The first aim was to optimize and evaluate the performance of the STS detection algorithm proposed by Adamowicz et al. on the different sensor positions lower back, chest, and ear. For this, an optimization process of the hyper-parameters of the algorithm was done with a grid-search for each sensor position using the previously existing dataset containing young, healthy adults (Dataset 1). The performance of the positions was already good before optimization with the chest and ear position achieving an F1-score above 90 %, and the lower back above 97 %. However, the optimization process resulted in improved performance for the chest and ear position (F1-score 91.7 % and 95.1 %, respectively). On the other hand, the performance of the lower back position did not improve, as the parameter set of the algorithm was originally defined for this position. The lower back sensor with the default parameter set showed the best F1-score with 97.3 % among all positions and parameter sets. The enhanced performance of the chest and ear position with the optimized parameter sets resulted from an enhanced recall, indicating fewer FN. Overall, the optimization process showed, that the algorithm by Adamowicz et al. could be adapted to other sensor positions. Furthermore, by appropriately adjusting the parameter sets for each new position, the already good performance could be further enhanced.

However, the algorithm showed reduced precision with the optimized parameter sets, primarily due to a lack of performance within daily tasks resulting in falsely detected transitions. Therefore, an additional postprocessing step was developed. ML classifiers were trained using the detected STS transitions for every sensor position to reduced the FPs. This resulted in a precision of 100 % for lower back and ear position, making the algorithm more robust, especially for real-world measurements. After applying the classifiers, the ear position achieved the best performance with an F1-score of 98.9 %, indicating that this position could be used in further investigations.

The next step of this thesis involved validating the ear position of the two-step development of the algorithm by Adamowicz et al. using the previously existing Dataset 2 with older, healthy adults. This dataset included data from two different IMU systems: the NilsPod sensor and the Signia hearing aid sensor. The performance of these two systems on the algorithm was compared. Results showed a better performance of the NilsPod sensors compared to the Signia sensors, resulting in a higher F1-score of around 5 % for both sensors, respectively. This could be attributed to the fact, that the grid-search and the training of the classifier were done based on data from a right NilsPod sensor. Additionally, the references were not reliable for the Signia sensors, resulting in an overall lower performance compared to the NilsPods. The lower performance of all four sensors in detecting STS transitions could be due to the development of the algorithm with the two-step approach using data from young, healthy participants.

Furthermore, the performance of the algorithm was validated on Dataset 3, which comprised older, frail participants. A study was conducted to assess data from these participants. During this study, participants were required to perform several STS transitions on two separate days. These transitions were carried out as part of the TUG test and were integrated into different small tasks. The performance of the algorithm was even lower within this Dataset 3 compared to the performance on Dataset 2. This could attributed to the fact that the STS transitions performed by the frail participants differ even more from those of the young, healthy ones compared to the older, healthy participants.

The last aim of this thesis was to assess the frailty status with the automatically detected STS transitions. For this purpose, all (frail) participants from Dataset 3 and all (frail and non-frail) participants from Dataset 2 were selected to train a classifier for frailty prediction. The trained classifier showed good performance in predicting frailty, achieving an F1-score of 71.4 %. However, due to the limited number of participants and the combination of two different datasets for frailty prediction, these results have to be treated with caution. A graphic representation of the most important features underlines the results of the classifier and demonstrated that there are differences in performing the STS transitions between the non-frail and frail groups. Therefore, further investigations should be done to achieve a reliable classifier for frailty prediction using more data for training the classifier.

The developed two-step approach showed lower performance on Dataset 2 and Dataset 3, which comprised different populations compared to Dataset 1. This suggests that the two-step approach was overfitted to the movement patterns of young, healthy adults, limiting its applicability to other populations. Therefore, in a further step, the grid-search for optimizing the parameter set of the algorithm and the training of the ML classifier should be done including a wider range of populations with participants from different age groups and health status to get as much motion patterns for STS transitions as possible and to make the algorithm more robust for different populations.

The approach of frailty prediction in this thesis showed, that there are differences in the transitions of non-frail and frail participants. However, to develop a robust classifier, a larger number of participants is needed. It is recommended to conduct a study with participants representing all three stages of frailty: non-frail, pre-frail and frail.

Furthermore, it is recommended to conduct a real-life study to further investigate frailty prediction in natural settings, potentially extending the study duration to assess the decline in frailty over time and observe any differences in movement patterns. In these subsequent studies, the frailty status of participants can be assessed using the Fried Frailty Index as the gold standard, enabling a comparison with the frailty status determined by the classifier.

To enhance the planning of the study and to ensure reliable results, it would be beneficial to recruit a larger number of participants who have to perform fewer STS transitions to ensure that all can successfully complete the tasks. Despite these limitations, this thesis showed promising results in predicting the frailty status based on detected STS transitions using ear-worn IMUs. These sensors offer an unobtrusive placement suitable for everyday life and can be integrated into hearing aids or earbuds.

Appendix A

Study description

Study design:

"Evaluation der Erkennung von Gebrechlichkeit über Aufstehen in verschiedenen Alltagssituationen"

1. General Information

1.1. Purpose

The aim of this study is to collect motion data from hearing aids for the development and evaluation of an algorithm for assessing frailty based on sit-to-stand transitions.

The level of frailty has an impact on the sit-to-stand movement, which is already being investigated in clinical tests. Frailty is a degenerative process, and therefore, information about the frailty level can be drawn from standing up in daily living, enabling early detection of this process.

In a previously conducted study, an algorithm for detecting sit-to-stand transfers using different sensor positions was developed and evaluated on young, healthy adults.

The aim of this study is to transfer this existing algorithm to older adults and to assess the level of frailty using the motion data. This data will be obtained using inertial measurement units integrated in hearing aids (accelerometer and gyroscope) worn on the ear.

1.2. Participants & Requirements

- 30 Participants
- Inclusion criteria:
  (1) Age ≥ 60 years
  (2) Ability to concentrate for 60 – 90 minutes
  (3) Ability to stand up independently
  (4) Willingness and ability to participate
- exclusion criteria
  (1) Not able to stand um 15 times
  (2) care level 5

1.3 Date & Location

- Pilot study: 26th May 2023
- Study: 8th May – 17th May 2023
- Duration: 60 min on day 1, 30 min on day 2
- Location: Haus Fronmüller, Fronmüllerstraße 129, 90763 Fürth

1.3. Sensor system

The sensors are ear-worn intertial measurement units (accelerometer and gyroscope). The sensors are integrated into a hearing aid housing and do not affect the hearing ability of the participants. The transitions were tracked using an app.

2. Study

Day 1:

Figure 1: Station with memory as task

Figure 2: Station with puzzle as task

Figure 3: Station with building bricks as task

Figure 4: Station with cross word as task

Figure 5: Picture to be described within a task from a Video, min 0:11[1]

Day 2:

Figure 6: Small tasks within transitions on day two

[1] https://www.youtube.com/watch?v=NeMB6XFi9M0

Zubehör für Studie

Unterlagen pro Probanden:

- Studienprotokoll (Tag 1 und 2)
- Datenschutzbestimmung + Einverständniserklärung
- Fragebögen für Probanden
    - Fried Frailty Phenotype
    - SARC-F
    - CFS (vorab an Personal)
    - ASTS (2x)
    -

Material für Aufnahme

- Handys (4) + Ladekabel + USB-C Kabel
- Hörgeräte
- Domes (versch. Größen)
- Receiver (versch. Größen)
- Batterien
-

Material für Aufgaben

- Memory
- Bausteine
- Kreuzworträtsel + Stift
- Puzzle
- Fineliner
- Bälle
-

Sonstiges

- Skalen der Fragebögen
- Handkraftmessgerät
- Klebeband
- Zollstock
- Laptop + Ladekabel
- Ordner für Unterlagen

Studienprotokoll – TAG 1		
VPN:		Datum:
Versuchsleitung:		

VORBEREITUNG
☐ Fragebögen bereitgelegt + beschriftet
☐ Fragebogen CFS von Pflegekräften ausgefüllt vorliegend (+ Einverständniserklärung)
☐ Sensor-ID, Handy-ID und VPN in Tabelle eingetragen
☐ Studienhandy aufgeladen, neue Batterie für Hörgerät
☐ Stationen mit Aufgaben vorbereitet: 5 Tische mit Stühlen (Kreuzworträtsel + Stift, Bausteine, Memory, Videos auf Laptop, Puzzle)
☐ Stuhl und Streckenmarkierung für TUG vorbereitet (Stuhlhöhe ca. 46 cm, 3 m hin- und zurück)
☐ Strecke für SPPB abgemessen und markiert (4 m)
☐ Strecke für Fried Frailty abgemessen und markiert (4,78 m) + Handkraftmessgerät bereit liegen

Start (Uhrzeit): _ _ : _ _	
☐ Begrüßung	☐ Ablauf erklärt
FRAGEBÖGEN	
☐ CFS (vorab von Pflegepersonal ausgefüllt)	*Punktzahl:* ___________
☐ SARC-F	*Punktzahl:* ___________
☐ ASTS	*Punktzahl:* ___________
☐ Fried Frailty	*Punktzahl:* ___________

SENSOR ANBRINGEN	*Label: „Nothing*
☐ Hörgerätsensor: Verbinden in App, **Title Messung: VPN + Tag 1**, DataRate: 50	
☐ Start Messung	
☐ Hörgerät anbringen (rechtes Ohr) + Sitz überprüfen	

TUG	*Label: „Walking“ jeweils für einen Durchgang, danach: „Nothing“*
Durchführung des TUG möglich? ☐ ja ☐ nein Gehilfe benötigt? ☐ ja ☐ nein Wenn ja, folgende Gehhilfe: ________________ 0 = keine, 1 = 1 Stock, 2 = 2 Stöcke, 3 = 4-Punktstock, 4 = 1 Krücke, 5 = 2 Krücken, 6 = Rollator, 7 = Gehgestell, 8 = Rollstuhl	Dauer: ________________ Auswertung: ____________________________ (Interpretation siehe Dokument *„TUG“)*

Natürliches Aufstehen	*Label VOR jedem Aufstehen: „Sitting“, dazwischen: „Nodding“*
Durchgang 1 ☐ Memory (6 Paare) ☐ Bausteine (Dorf bauen) ☐ Video anschauen (Frage: Sieht man in dem Video verschiedene Tiere oder Landschaft?) ☐ Puzzle (Variante 1) ☐ Kreuzworträtsel (2 min Timer)	*Anzahl absolvierter Aufgaben:* ________
Durchgang 2 ☐ Memory 2 (6 Paare) ☐ Bausteine (hohen Turm bauen) ☐ Video anschauen (Frage: Sieht man in dem Video das Meer oder Bäume?) ☐ Puzzle (Variante 2) ☐ Kreuzworträtsel (2 min Timer)	*Anzahl absolvierter Aufgaben:* ________

SPPB	*Label: „Running“* *bei 5x STS: „Talking“* *nach Beenden: „Nothing“*
1. Balance-Test	
Geschlossener Stand	*Dauer:* __________ *(max. 10 sek)* *Punktzahl*: ___________
Semi-Tandem Stand	*Dauer:* __________ *(max. 10 sek)* *Punktzahl*: ___________
Tandem Stand	*Dauer:* __________ *(max. 10 sek)* *Punktzahl*: ___________
Auswertung: 10 Sek. – 1 Pkt., < 10 Sek. – 0 Pkt.	
2. Gehgeschwindigkeit	
4 m (2 Versuche: schnellerer zählt, Hilfsmittel erlaubt) Benötigtes Hilfsmittel: _________________	*Versuch 1:* __________ *(max. 10 sek)* *Versuch 2:* __________ *(max. 10 sek)* *Punktzahl*: ___________
Auswertung: *4 Pkt. ≤ 4.8 Sek, 3 Pkt. 4.8 – 6.2 Sek., 2 Pkt. 6.2 – 8.7 Sek., 1 Pkt. > 8.7 Sek., 0 Pkt. Nicht möglich*	
3. Chair Stand Test	
1x Aufstehen möglich? ☐ ja ☐ nein 5x Aufstehen (ohne Armlehnen, Arme vor Brust verschränkt)	*Dauer:* __________ *(max. 10 sek)* *Punktzahl*: ___________
Auswertung: *4 Pkt. < 11.2 Sek., 3 Pkt. 11.2 – 13.6 Sek., 2 Pkt. 13.7 – 16.6 Sek., 1 Pkt. > 16.6 Sek., 0 Pkt. > 60 Sek. oder nicht möglich*	
Summe Balance Test: ________ *Summe 4-m Gehtest:* _________ *Summe CST:* _________ *Gesamtpunktzahl SPPB:* ___________	

NACHBEREITUNG
☐ Sensor abnehmen ☐ Messung beenden (Hörgeräte-App) ☐ Daten des Sensors sichern ☐ alle Punktzahlen für Tests eingetragen ☐ Checkliste in Studienordner abgeheftet + Studienprotokoll abgeheftet Für Langzeitmessung: ☐ Sensor-ID + Handy-ID + VPN in Tabelle eintragen ☐ Sensor mit Handy für Langzeitmessung verbinden ☐ Sensor anbringen + Pfleger/Proband einweisen ☐ Messung starten

Anmerkungen:

Studienprotokoll – TAG 2		
VPN:		Datum:
Versuchsleitung:		

VORBEREITUNG
☐ Fragebogen bereitgelegt + beschriftet ☐ Sensor-ID, Handy-ID und VPN in Tabelle eintragen ☐ Studienhandy aufgeladen, neue Batterie für Hörgerät ☐ 3 verschiedene Stuhltypen vorbereiten + kleine Aufgaben in Reichweite (Stifte und Bälle) ☐ Stuhl und Streckenmarkierung für TUG vorbereitet (Stuhlhöhe ca. 46 cm, 3 m hin- und zurück) ☐ Strecke für SPPB abgemessen und markiert (4 m)

Start (Uhrzeit): _ _ : _ _	
☐ Begrüßung	☐ Ablauf erklärt
FRAGEBOGEN	
☐ ASTS	*Punktzahl:* __________

SENSOR ANBRINGEN	*Label: „Nothing“*
☐ Hörgerätsensor: Verbinden, **Title Messung: VPN + Tag 2**, DataRate: 50 ☐ Start Messung ☐ Hörgerät anbringen (rechtes Ohr) + Sitz überprüfen	

<table>
<tr><th>TUG</th><th>Label: „Walking“
nach beiden Durchgängen: „Nothing“</th></tr>
<tr><td>Durchführung des TUG möglich?
☐ ja ☐ nein

Gehilfe benötigt?
☐ ja ☐ nein
Wenn ja, folgende Gehhilfe: ________________
0 = keine, 1 = 1 Stock, 2 = 2 Stöcke,
3 = 4-Punktstock, 4 = 1 Krücke, 5 = 2 Krücken,
6 = Rollator, 7 = Gehgestell, 8 = Rollstuhl</td><td>Dauer: ________________

Auswertung: ____________________________
(Interpretation siehe Dokument „TUG“)</td></tr>
</table>

<table>
<tr><th colspan="2">5 x Aufstehen</th><th>Label VOR jedem Aufstehen: „Sitting“, dazwischen: „Nodding“</th></tr>
<tr><td>Reihenfolge: _____</td><td>Normalhöhe, ohne Armlehne</td><td>Anzahl absolvierter Transitionen: ________</td></tr>
<tr><td>Reihenfolge: _____</td><td>Sofa</td><td>Anzahl absolvierter Transitionen: ________</td></tr>
<tr><td>Reihenfolge: _____</td><td>Normalhöhe, mit Armlehne</td><td>Anzahl absolvierter Transitionen: ________</td></tr>
</table>

<table>
<tr><th>SPPB</th><th>Label: „Running“
bei 5x STS: „Talking“
nach Beenden: „Nothing“</th></tr>
<tr><td colspan="2">1. Balance-Test</td></tr>
<tr><td>Geschlossener Stand</td><td>Dauer: _________ (max. 10 sek)
Punktzahl: ___________</td></tr>
<tr><td>Semi-Tandem Stand</td><td>Dauer: _________ (max. 10 sek)
Punktzahl: ___________</td></tr>
<tr><td>Tandem Stand</td><td>Dauer: _________ (max. 10 sek)
Punktzahl: ___________</td></tr>
<tr><td colspan="2">Auswertung: 10 Sek. – 1 Pkt., < 10 Sek. – 0 Pkt.</td></tr>
</table>

2. Gehgeschwindigkeit	
4 m (2 Versuche: schnellerer zählt, Hilfsmittel erlaubt) Benötigtes Hilfsmittel: ____________	*Versuch 1:* ________ *(max. 10 sek)* *Versuch 2:* ________ *(max. 10 sek)* *Punktzahl*: ________
Auswertung: *4 Pkt. ≤ 4.8 Sek, 3 Pkt. 4.8 – 6.2 Sek., 2 Pkt. 6.2 – 8.7 Sek., 1 Pkt. > 8.7 Sek., 0 Pkt. Nicht möglich*	
3. Chair Stand Test	
1x Aufstehen möglich? ☐ ja ☐ nein 5x Aufstehen (ohne Armlehnen, Arme vor Brust verschränkt)	*Dauer:* ________ *(max. 10 sek)* *Punktzahl*: ________
Auswertung: *4 Pkt. < 11.2 Sek., 3 Pkt. 11.2 – 13.6 Sek., 2 Pkt. 13.7 – 16.6 Sek., 1 Pkt. > 16.6 Sek., 0 Pkt. > 60 Sek. oder nicht möglich*	
Summe Balance Test: ________ *Summe 4-m Gehtest:* ________ *Summe CST:* ________ *Gesamtpunktzahl SPPB:* ________	

NACHBEREITUNG
☐ Sensor abnehmen ☐ Messung beenden (Hörgeräte-App) ☐ Daten des Sensors sichern ☐ alle Punktzahlen für Tests eingetragen ☐ Checkliste in Studienordner abgeheftet + Studienprotokoll abgeheftet

Anmerkungen:

List of Figures

List of Tables

Bibliography

[Abd21] Vahid Abdollah, Tarek N. Dief, John Ralston, Chester Ho, and Hossein Rouhani. "Investigating the validity of a single tri-axial accelerometer mounted on the head for monitoring the activities of daily living and the timed-up and go test". In: *Gait & Posture* 90 (2021), pp. 137–140. DOI: 10.1016/j.gaitpost.2021.08.020.

[Ada20a] Lukas Adamowicz, F. Isik Karahanoglu, Christopher Cicalo, Hao Zhang, Charmaine Demanuele, Mar Santamaria, Xuemei Cai, and Shyamal Patel. "Assessment of sit-to-stand transfers during daily life using an accelerometer on the lower back". In: *Sensors* 20.22 (2020). DOI: 10.3390/s20226618.

[Ada20b] Lukas Adamowicz and Shyamal Patel. "Sit2StandPy: An open-source python package for detecting and quantifying sit-to-stand transitions using an accelerometer on the lower back". In: *Journal of Open Source Software* 5 (2020), p. 2249. DOI: 10.21105/joss.02449.

[Ahm13] Norhafizan Ahmad, Raja Ariffin Raja Ghazilla, Nazirah M. Khairi, and Vijayabaskar Kasi. "Reviews on various inertial measurement unit (IMU) sensor applications". In: *International Journal of Signal Processing Systems* (2013), pp. 256–262. DOI: 10.12720/ijsps.1.2.256-262.

[Ale91] Neil B. Alexander, Albert B. Schultz, and David N. Warwick. "Rising from a chair: effects of age and functional ability on performance biomechanics". In: *Journal of Gerontology* 46.3 (1991), pp. M91–M98. DOI: 10.1093/geronj/46.3.m91.

[Ata11] Louis Atallah, Benny Lo, Rachel King, and Yang Guang-Zhong. "Sensor positioning for activity recognition using wearable accelerometers". In: *IEEE Transactions on Biomedical Circuits and Systems* 5.4 (2011), pp. 320–329. DOI: 10.1109/TBCAS.2011.2160540.

[Atr20] Arash Atrsaei, Farzin Dadashi, Clint Hansen, Elke Warmerdam, Benoît Mariani, Walter Maetzler, and Kamiar Aminian. "Postural transitions detection and characterization in healthy and patient populations using a single waist sensor". In: *Journal of NeuroEngineering and Rehabilitation* 17.1 (2020), p. 70. DOI: 10.1186/s12984-020-00692-4.

[Ben21] Petra Benzinger, Annette Eidam, and Jürgen M. Bauer. "Klinische Bedeutung der Erfassung von Frailty". In: *Zeitschrift für Gerontologie und Geriatrie* 54.3 (2021), pp. 285–296. DOI: 10.1007/s00391-021-01873-z.

[Ber04] M. Bernardi, A. Rosponi, V. Castellano, A. Rodio, J. M. Traballesi, A. S. Delussu, and M. Marchetti. "Determinants of sit-to-stand capability in the motor impaired elderly". In: *Journal of Electromyography and Kinesiology* 14 (2004), pp. 401–410. DOI: 10.1016/j.jelekin.2003.09.001.

[Bis09] Christopher M. Bishop. *Pattern recognition and machine learning*. 9. (corrected at 8th printing). Information science and statistics. Springer New York, 2009.

[Bon17] Giuseppe Bonaccorso. *Machine learning algorithms: A reference guide to popular algorithms for data science and machine learning*. Birmingham and Mumbai: Packt Publishing Ltd., 2017.

[Bra18] Tobias Braun, Christian Grüneberg, and Christian Thiel. "Deutsche Übersetzung, interkulturelle Adaptation und diagnostische Testgenauigkeit von drei Frailty Screening Fragebögen : PRISMA-7, FRAIL Skala und Groningen Frailty Indicator". In: *Zeitschrift fur Gerontologie und Geriatrie* 51.3 (2018), pp. 282–292. DOI: 10.1007/s00391-017-1295-2.

[Bre01] Leo Breiman. "Random forests". In: *Machine Learning* 45.1 (2001), pp. 5–32. DOI: 10.1023/A:1010933404324.

[Bro10] Nefertiti A. Brown and Michael E. Zenilman. "The impact of frailty in the elderly on the outcome of surgery in the aged". In: *Advances in Surgery* 44 (2010), pp. 229–249. ISSN: 0065-3411. DOI: 10.1016/j.yasu.2010.05.014.

[Bur20] Clara Piris Burgos, Lea Gartner, Miguel A. Gonzalez Ballester, Jerome Noailly, Fabian Stocker, Martin Schonfelder, Tim Adams, and Simone Tassani. "In-Ear accelerometer-based sensor for gait classification". In: *IEEE Sensors Journal* 20.21 (2020), pp. 12895–12902. DOI: 10.1109/JSEN.2020.3002589.

[But16] Brian J. Buta, Jeremy D. Walston, Job G. Godino, Minsun Park, Rita R. Kalyani, Qian-Li Xue, Karen Bandeen-Roche, and Ravi Varadhan. “Frailty assessment instruments: Systematic characterization of the uses and contexts of highly-cited instruments”. In: *Ageing Research Reviews* 26 (2016), pp. 53–61. DOI: 10.1016/j.arr.2015.12.003.

[Cam89] A. John Campbell, Michael J. Borrie, and Georg F. Spears. “Risk factors for falls in a community-based prospective study of people 70 years and older”. In: *Journal of Gerontology* 44.4 (1989), pp. M112–M117. DOI: 10.1093/geronj/44.4.m112.

[Chr18] Maximilian Christ, Nils Braun, Julius Neuffer, and Andreas W. Kempa-Liehr. “Time series feature extraction on basis of scalable hypothesis tests (tsfresh – a python package)”. In: *Neurocomputing* 307 (2018), pp. 72–77. DOI: 10.1016/j.neucom.2018.03.067.

[Cob20] Antonio Cobo, Elena Villalba-Mora, Rodrigo Pérez-Rodríguez, Xavier Ferre, Walter Escalante, Cristian Moral, and Leocadio Rodriguez-Mañas. “Automatic and real-time computation of the 30-seconds chair-stand test without professional supervision for community-dwelling older adults”. In: *Sensors* 20.20 (2020). DOI: 10.3390/s20205813.

[Csu85] Maryellen Csuka and Daniel J. McCarty. “Simple method for measurement of lower extremity muscle strength”. In: *The American Journal of Medicine* 78.1 (1985), pp. 77–81. ISSN: 0002-9343. DOI: 10.1016/0002-9343(85)90465-6.

[Dal02] Claudia Dalbert. *ASTS - Aktuelle Stimmungsskala*. ZPID (Leibniz Institute for Psychology Information) - Testarchiv, 2002. DOI: 10.23668/psycharchives.327. URL: https://www.psycharchives.org/handle/20.500.12034/340. Accessed: 2023-07-07.

[Fer94] F. Ferraris, I. Gorini, U. Grimaldi, and M. Parvis. “Calibration of three-axial rate gyros without angular velocity standards”. In: *Sensors and Actuators A: Physical* 42 (1994), pp. 446–449. DOI: 10.1016/0924-4247(94)80031-6.

[Fri01] Linda P. Fried, Catherine M. Tangen, Jeremy Walston, Anne B. Newman, Calvin Hirsch, John Gottdiener, Teresa Seeman, Russell Tracy, Willem J. Kop, Gregory Burke, and Mary Ann McBurnie. “Frailty in older adults: evidence for a phenotype". In: *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 56.3 (2001), pp. M146–M157. DOI: 10.1093/gerona/56.3.M146.

[Gan11] R. Ganea, A. Paraschiv-Ionescu, C. Büla, S. Rochat, and K. Aminian. "Multiparametric evaluation of sit-to-stand and stand-to-sit transitions in elderly people". In: *Medical Engineering & Physics* 33.9 (2011), pp. 1086–1093. DOI: 10.1016/j.medengphy.2011.04.015.

[Grä04] Kerstin Gräfe, Stephan Zipfel, Wolfgang Herzog, and Bernd Löwe. "Screening psychischer Störungen mit dem "Gesundheitsfragebogen für Patienten (PHQ-D)"". In: *Diagnostica* 50.4 (2004), pp. 171–181. DOI: 10.1026/0012-1924.50.4.171.

[Gre14] Barry R. Greene, Emer P. Doheny, Aisling O'Halloran, and Rose Anne Kenny. "Frailty status can be accurately assessed using inertial sensors and the TUG test". In: *Age and Ageing* 43.3 (2014), pp. 406–411. DOI: 10.1093/ageing/aft176.

[Gur94] Jack M. Guralnik, Eleanor M. Simonsick, Luigi Ferrucci, Robert J. Glynn, Lisa F. Berkman, Dan G. Blazer, Paul A. Scherr, and Robert B. Wallace. "A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission". In: *Journal of Gerontology* 49.2 (1994), pp. M85–M94. DOI: 10.1093/geronj/49.2.m85.

[Hap21] Johanna Happold, Robert Richer, Arne Kuderle, Heiko Gabner, Jochen Klucken, Bjoern M. Eskofier, and Felix Kluge. "Evaluation of orthostatic reactions in real-world environments using wearable sensors". In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 2021 (2021), pp. 6987–6990. DOI: 10.1109/EMBC46164.2021.9630842.

[Hel19] Sandra Hellmers, Sebastian Fudickar, Sandra Lau, Lena Elgert, Rebecca Diekmann, Jürgen M. Bauer, and Andreas Hein. "Measurement of the chair rise performance of older people based on force plates and IMUs". In: *Sensors* 19.6 (2019), p. 1370. DOI: 10.3390/s19061370.

[Hug96] M. A. Hughes and M. L. Schenkman. "Chair rise strategy in the functionally impaired elderly". In: *Journal of Rehabilitation Research and Development* 33.4 (1996), pp. 409–412.

[Jon99] C. Jessie Jones, Roberta E. Rikli, and William C. Beam. "A 30-s chair-stand test as a measure of lower body strength in community-residing older adults". In: *Research Quarterly for Exercise and Sport* 70.2 (1999), pp. 113–119. DOI: 10.1080/02701367.1999.10608028.

[Kha22] Nasim B. Khan, Nolwazi Mthembu, Aishwarya Narothan, Zamahlase Sibisi, and Qiniso Vilane. "Health sciences students' perception of the communicative impacts of face coverings during the COVID-19 pandemic at a South African University". In: *The South African Journal of Communication Disorders* 69.2 (2022), e1–e14. DOI: 10.4102/sajcd.v69i2.890.

[Khe99] M.M. Khemlani, J.H. Carr, and W.J. Crosbie. "Muscle synergies and joint linkages in sit-to-stand under two initial foot positions". In: *Clinical Biomechanics* 14 (1999), pp. 236–246. DOI: 10.1016/s0268-0033(98)00072-2.

[Kia97] K. Kiani, C. J. Snijders, and E. S. Gelsema. "Computerized analysis of daily life motor activity for ambulatory monitoring". In: *Technology and Health Care* 5.4 (1997), pp. 307–318. DOI: 10.3233/THC-1997-5404.

[Küd23] Arne Küderle, Robert Richer, Raul C. Sîmpetru, and Bjoern M. Eskofier. "tpcp: Tiny pipelines for complex problems - A set of framework independent helpers for algorithms development and evaluation". In: *Journal of Open Source Software* 8 (2023), p. 4953. DOI: 10.21105/joss.04953.

[Lac12] Alethea Lacas and Kenneth Rockwood. "Frailty in primary care: a review of its conceptualization and implications for practice". In: *BMC Medicine* 10 (2012), p. 4. DOI: 10.1186/1741-7015-10-4.

[Lep18] Kevin Lepetit, Khalil Ben Mansour, Sofiane Boudaoud, Kiyoka Kinugawa-Bourron, and Frédéric Marin. "Evaluation of the kinetic energy of the torso by magneto-inertial measurement unit during the sit-to-stand movement". In: *Journal of Biomechanics* 67 (2018), pp. 172–176. DOI: 10.1016/j.jbiomech.2017.11.028.

[Lep19] Kevin Lepetit, Khalil Ben Mansour, Adrien Letocart, Sofiane Boudaoud, Kiyoka Kinugawa, Jean-François Grosset, and Frédéric Marin. "Optimized scoring tool to quantify the functional performance during the sit-to-stand transition with a magneto-inertial measurement unit". In: *Clinical Biomechanics* 69 (2019), pp. 109–114. DOI: 10.1016/j.clinbiomech.2019.07.012.

[Mag07] I. Maglogiannis, K. Karpouzis, B.A. Wallace, and J. Soldatos, eds. *Emerging artificial intelligence applications in computer engineering: Real word AI systems with applications in eHealth, HCI, information retrieval and pervasive technologies*. Vol. 160. Frontiers in artificial intelligence and applications. Amsterdam: IOS Press, 2007.

[Mal13] Theodore K. Malmstrom and John E. Morley. "SARC-F: a simple questionnaire to rapidly diagnose sarcopenia". In: *Journal of the American Medical Directors Association* 14.8 (2013), pp. 531–532. DOI: 10.1016/j.jamda.2013.05.018.

[Mar19] Uriel Martinez-Hernandez and Abbas A. Dehghani-Sanij. "Probabilistic identification of sit-to-stand and stand-to-sit with a wearable sensor". In: *Pattern Recognition Letters* 118 (2019), pp. 32–41. DOI: 10.1016/j.patrec.2018.03.020.

[Mil13a] Nora Millor, Pablo Lecumberri, Marisol Gomez, Alicia Martinez-Ramirez, Leocadio Rodriguez-Manas, Francisco José Garcia-Garcia, and Mikel Izquierdo. "Automatic evaluation of the 30-s chair stand test using inertial/magnetic-based technology in an older prefrail population". In: *IEEE Journal of Biomedical and Health Informatics* 17.4 (2013), pp. 820–827. DOI: 10.1109/JBHI.2013.2238243.

[Mil13b] Nora Millor, Pablo Lecumberri, Marisol Gómez, Alicia Martínez-Ramírez, and Mikel Izquierdo. "An evaluation of the 30-s chair stand test in older adults: frailty detection based on kinematic parameters from a single inertial unit". In: *Journal of Neuroengineering and Rehabilitation* 10 (2013), p. 86. DOI: 10.1186/1743-0003-10-86.

[Mil14a] Nora Millor, Pablo Lecumberri, Marisol Gomez, Alicia Martinez-Ramirez, and Mikel Izquierdo. "Kinematic parameters to evaluate functional performance of sit-to-stand and stand-to-sit transitions using motion sensor devices: a systematic review". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22.5 (2014), pp. 926–936. DOI: 10.1109/TNSRE.2014.2331895.

[Mil14b] Nora Millor, Pablo Lecumberri, Alicia Martinez-Ramirez, Jon Martinikorena, Marisol Gómez, and Mikel Izquierdo. "Instrumented 30-s Chair Stand Test: evaluation of an exercise program in frail nonagenarians". In: *IWBBIO* (2014), pp. 1618–1629. URL: http://iwbbio.ugr.es/2014/papers/IWBBIO_2014_paper_179.pdf. Accessed: 2023-07-07.

[Mil17] Nora Millor, Pablo Lecumberri, Marisol Gomez, Alicia Martinez, Jon Martinikorena, L. Rodriguez-Manas, F. J. Garcia-Garcia, and Mikel Izquierdo. "Gait velocity and chair sit-stand-sit performance improves current frailty-status identification". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25.11 (2017), pp. 2018–2025. DOI: 10.1109/TNSRE.2017.2699124.

[Min18] Chulhong Min, Akhil Mathur, and Fahim Kawsar. "Exploring audio and kinetic sensing on earable devices". In: *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications*. Ed. by Kaspar Jansen and Ashwin Ashok. New York, NY, USA: ACM, 2018, pp. 5–10. ISBN: 9781450358422. DOI: 10.1145/3211960.3211970.

[Mor12] J. E. Morley, T. K. Malmstrom, and D. K. Miller. "A simple frailty questionnaire (FRAIL) predicts outcomes in middle aged African Americans". In: *The Journal of Nutrition, Health & Aging* 16.7 (2012), pp. 601–608. DOI: 10.1007/s12603-012-0084-2.

[Mou00] France Moureya, Alexandre Grishinb, Philippe d'Athisc, Thierry Pozzod, and Paul Stapleyd. "Standing up from a chair as a dynamic equilibrium task: A comparison between young and elderlys". In: *The Journals of Gerontology: Series A* 55 (2000), pp. 425–431. DOI: 10.1093/gerona/55.9.b425.

[Mur21] Paul Murphy, Dearbhail Lewis, and Gerard J. Gormley. "Caring from behind the face mask in healthcare: Learning from the dramatic arts". In: *Perspectives on Medical Education* 10.6 (2021), pp. 352–355. DOI: 10.1007/s40037-021-00691-8.

[Mur98] Sreerama K. Murthy. "Automatic construction of decision trees from data: A multi-disciplinary survey". In: *Data Mining and Knowledge Discovery* 2.4 (1998), pp. 345–389. DOI: 10.1023/A:1009744630224.

[Naj02] Bijan Najafi, Kamiar Aminian, François Loew, Yves Blanc, and Philippe A. Robert. "Measurement of stand-sit and sit-stand transitions using a miniature gyroscope and its application in fall risk evaluation in the elderly". In: *IEEE Transactions on Biomedical Engineering* 49.8 (2002), pp. 843–851. DOI: 10.1109/TBME.2002.800763.

[Ngu15] Hung P. Nguyen, Fouaz Ayachi, Catherine Lavigne-Pelletier, Margaux Blamoutier, Fariborz Rahimi, Patrick Boissy, Mandar Jog, and Christian Duval. "Auto detection and segmentation of physical activities during a Timed-Up-and-Go (TUG) task in healthy older adults using multiple inertial sensors". In: *Journal of neuroengineering and rehabilitation* 12 (2015), p. 36. DOI: 10.1186/s12984-015-0026-4.

[Nie83] Heinrich Niemann. *Klassifikation von Mustern*. Springer Verlag, 1983. DOI: 10.1007/978-3-642-47517-7.

[Nuz86] Sharon Nuzik, Robert Lamb, Ann VanSant, and Susanne Hirt. "Sit-to-stand movement pattern. A kinematic study". In: *Physical Therapy* 66.11 (1986), pp. 1708–1713. DOI: 10.1093/ptj/66.11.1708.

[Oll22] Malte Ollenschläger, Arne Küderle, Wolfgang Mehringer, Ann-Kristin Seifer, Jürgen Winkler, Heiko Gaßner, Felix Kluge, and Bjoern M. Eskofier. "MaD GUI: An open-source Python package for annotation and analysis of time-series data". In: *Sensors* 22.15 (2022). ISSN: 1424-8220. DOI: 10.3390/s22155849.

[Pai94] Yi-Chung Pai and Wynne A. Lee. "Effect of a terminal constraint on control of balance during sit-to-stand". In: *Journal of Motor Behavior* 26 (1994), pp. 247–256. DOI: 10.1080/00222895.1994.9941680.

[Par17] Saman Parvaneh, Jane Mohler, Nima Toosizadeh, Gurtej Singh Grewal, and Bijan Najafi. "Postural transitions during activities of daily living could identify frailty status: application of wearable technology to identify frailty during unsupervised condition". In: *Gerontology* 63.5 (2017), pp. 479–487. DOI: 10.1159/000460292.

[Ped11] 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". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[Pic19] Chris G. Pickford, Andrew H Findlow, Andy Kerr, Matthew Banger, Alexandra M. Clarke-Cornwell, Kristen L. Hollands, Terry Quinn, and Malcolm H. Granat. "Quantifying sit-to-stand and stand-to-sit transitions in free-living environments using the activPAL thigh-worn activity monitor". In: *Gait & Posture* 73 (2019), pp. 140–146. DOI: 10.1016/j.gaitpost.2019.07.126.

[Pod91] Diane Podsiadlo and Sandra Richardson. "The timed "Up & Go": a test of basic functional mobility for frail elderly persons". In: *Journal of the American Geriatrics Society* 39.2 (1991), pp. 142–148. DOI: 10.1111/j.1532-5415.1991.tb01616.x.

[Reg15] G. Ruben H. Regterschot, Wei Zhang, Heribert Baldus, Martin Stevens, and Wiebren Zijlstra. "Sensor-based monitoring of sit-to-stand performance is indicative of objective and self-reported aspects of functional status in older adults". In: *Gait & Posture* 41.4 (2015), pp. 935–940. DOI: 10.1016/j.gaitpost.2015.03.350.

[Ric21] Robert Richer, Arne Küderle, Martin Ullrich, Nicolas Rohleder, and Bjoern M. Eskofier. "BioPsyKit: A Python package for the analysis of biopsychological data". In: *Journal of Open Source Software* 6.66 (2021), p. 3702. DOI: 10.21105/joss.03702.

[Roc05] Kenneth Rockwood, Xiaowei Song, Chris MacKnight, Howard Bergman, David B. Hogan, Ian McDowell, and Arnold Mitnitski. "A global clinical measure of fitness and frailty in elderly people". In: *CMAJ : Canadian Medical Association Journal* 173.5 (2005), pp. 489–495. DOI: 10.1503/cmaj.050051.

[Roc07] Kenneth Rockwood and Arnold Mitnitski. "Frailty in relation to the accumulation of deficits". In: *The Journals of Gerontology* 62.7 (2007), pp. 722–727. DOI: 10.1093/gerona/62.7.722.

[Roc11] Kenneth Rockwood and Arnold Mitnitski. "Frailty defined by deficit accumulation and geriatric medicine defined by frailty". In: *Clinics in Geriatric Medicine* 27.1 (2011), pp. 17–26. DOI: 10.1016/j.cger.2010.08.008.

[Roc99] Kenneth Rockwood, Karen Stadnyk, Chris MacKnight, Ian McDowell, Réjean Hébert, and David B Hogan. "A brief clinical instrument to classify frailty in elderly people". In: *The Lancet* 353.9148 (1999), pp. 205–206. DOI: 10.1016/S0140-6736(98)04402-X.

[Röd22] Tobias Röddiger, Christopher Clarke, Paula Breitling, Tim Schneegans, Haibin Zhao, Hans Gellersen, and Michael Beigl. “Sensing with Earables”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6.3 (2022), pp. 1–57. DOI: 10.1145/3550314.

[Rot18] Nils Roth, Christine F. Martindale, Bjoern M. Eskofier, Heiko Gaßner, Zacharias Kohl, and Jochen Klucken. “Synchronized sensor insoles for clinical gait analysis in home-monitoring applications”. In: *Current Directions in Biomedical Engineering* 4 (2018), pp. 433–437. DOI: 10.1515/cdbme-2018-0103.

[Sal07] Arash Salarian, Heike Russmann, Christian Wider, Pierre R. Burkhard, Franios J. G. Vingerhoets, and Kamiar Aminian. “Quantification of tremor and bradykinesia in Parkinson’s disease using a novel ambulatory monitoring system”. In: *IEEE transactions on Biomedical Engineering* 54.2 (2007), pp. 313–322. ISSN: 0018-9294. DOI: 10.1109/TBME.2006.886670.

[Sav13] George M. Savva, Orna A. Donoghue, Frances Horgan, Claire O’Regan, Hilary Cronin, and Rose Anne Kenny. “Using timed up-and-go to identify frail members of the older population”. In: *The Journals of Gerontology. Series A, Biological sciences and medical sciences* 68.4 (2013), pp. 441–446. DOI: 10.1093/gerona/gls190.

[Sei12] Christina Seimetz, Danica Tan, Riemann Katayama, and Thurmon Lockhar. “A comparison between methods of measuring postrual stability: force plates versus accelerometers”. In: *Biomed Sciences Instrumentation* 48 (2012), pp. 386–392.

[Shk21] Andrei M. Shkel and Yusheng Wang, eds. *Pedestrian inertial navigation with self-contained aiding*. IEEE Press Series on Sensors Ser. Newark: John Wiley & Sons Incorporated, 2021. DOI: 10.1002/9781119699910.

[Szt16] Timo Sztyler and Heiner Stuckenschmidt. “On-body localization of wearable devices: An investigation of position-aware activity recognition”. In: *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2016, pp. 1–9. DOI: 10.1109/PERCOM.2016.7456521.

[Tho92] Scott Thomas, J. Reading, and R. J. Shephard. “Revision of the physical activity readiness questionnaire (PAR-Q)”. In: *Canadian Journal of Sport Sciences* 17 (1992), pp. 338–345.

[Ull23] Martin Ullrich, Nils Roth, Arne Küderle, Robert Richer, Till Gladow, Heiko Gaßner, Franz Marxreiter, Jochen Klucken, Bjoern M. Eskofier, and Felix Kluge. "Fall risk prediction in parkinson's disease using real-world inertial sensor gait data". In: *IEEE Journal of Biomedical and Health Informatics* 27.1 (2023), pp. 319–328. DOI: 10.1109/JBHI.2022.3215921.

[van18] Rob C. van Lummel, Jordi Evers, Martijn Niessen, Peter J. Beek, and Jaap H. van Dieën. "Older adults with weaker muscle strength stand up from a sitting position with more dynamic trunk use". In: *Sensors (Basel, Switzerland)* 18.4 (2018). DOI: 10.3390/s18041235.

[Wal12] Jeremy D. Walston. "Sarcopenia in older adults". In: *Current Opinion in Rheumatology* 24.6 (2012), pp. 623–627. DOI: 10.1097/BOR.0b013e328358d59b.

[Zha12] He Zhao and Zheyao Wang. "Motion measurement using inertial sensors, ultrasonic sensors, and magnetometers with extended kalman filter for data fusion". In: *IEEE Sensors Journal* 12.5 (2012), pp. 943–953. DOI: 10.1109/JSEN.2011.2166066.

[Zha14] Wei Zhang, G. Ruben H. Regterschot, Fabian Wahle, Hilde Geraedts, Heribert Baldus, and Wiebren Zijlstra. "Chair rise transfer detection and analysis using a pendant sensor: An algorithm for fall risk assessment in older people". In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2014), pp. 1830–1834. DOI: 10.1109/EMBC.2014.6943965.

[Zha17] Wei Zhang, G. Ruben H. Regterschot, Hilde Geraedts, Heribert Baldus, and Wiebren Zijlstra. "Chair rise peak power in daily life measured with a pendant sensor associates with mobility, limitation in activities, and frailty in old people". In: *IEEE Journal of Biomedical and Health Informatics* 21.1 (2017), pp. 211–217. DOI: 10.1109/JBHI.2015.2501828.

Appendix B

Acronyms

AUC area under curve

IMU Inertial Measurement Unit

STS Sit-to-Stand

TUG timed up and go

30s CST 30 seconds chair stand test

PD Parkinson's disease

SVM support vector machine

ML Machine Learning

5xSTS five times sit-to-stand

CWT continuous wavelet transform

k-NN k-nearest neighbour

PCA principal component analysis

SPPB Short Physical Performance Battery

TP true positive

FP false positive

FN false negative

TN true negative

PAR-Q Physcial Activity Readiness Questionnaire

DT decision tree

SVM-lin support vector machine with linear basis function

SVM-rbf support vector machine with radial basis function

RF random forest

CV cross-validation

ECG electrocardiogramm

PFD physical frailty phenotype

CFS Clinical Frailty Scale

ASTS aktuelle Stimmungsskala

SARC-F Sarcopedia

FFP Fried Frailty Phenotype

PHQ-8 Patient Health Questionnaire

API application programming interface

HPC high-performance cluster

CoV coefficient of variation

SPARC spectral arc length

FAU Friedrich-Alexander-Universität Erlangen-Nürnberg