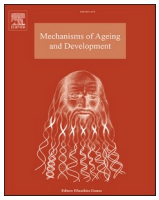
Mechanisms of Ageing and Development 228 (2025) 112114

Contents lists available at ScienceDirect

Mechanisms of Ageing and Development

journal homepage: www.elsevier.com/locate/mechagedev



Clinical data-driven classification of pre-frailty reveals sex-specific patterns – Data from the Berlin Aging Study II (BASE-II)

Jeff Didier [a], Sébastien De Landtsheer [a], Maria Pires Pacheco [a], Ali Kishk [a], Jochen G. Schneider [b,c,d], David Goldeck [e], Graham Pawelec [f], Dominik Spira [g], Ilja Demuth [g,h], Thomas Sauter [a,*]

[a] Systems Biology & Epigenetics Group, Department of Life Sciences and Medicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
[b] Medical Translational Research Group, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
[c] Department of Life Sciences and Medicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
[d] Department of Internal Medicine, Saarland University Hospital and Saarland University Faculty of Medicine, Homburg, Germany
[e] Department of Internal Medicine 2, University of Tübingen, Tübingen, Germany
[f] Department of Immunology, University of Tübingen, Tübingen, Germany
[g] Division of Lipid Metabolism, Department of Endocrinology and Metabolic Diseases, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Charité – Universitätsmedizin Berlin, Berlin, Germany
[h] Berlin Institute of Health at Charité – Universitätsmedizin Berlin, BCRT-Berlin Institute of Health Center for Regenerative Therapies, Berlin, Germany

ARTICLE INFO

Keywords:
Ageing
Frailty
Classification
Prognostic biomarkers
Machine-learning
BASE-II

ABSTRACT

Frailty is a geriatric condition with multidimensional consequences that strongly affect older adults' quality of life. The lack of a universal standard to describe, diagnose, and treat frailty further complicates this situation. Nowadays, multitudinous frailty assessment tools are applied depending on the regional and clinical context, adding complexity by increasing heterogeneity in the definition and characterization of frailty. Better insights into the causes and pathophysiology of frailty and its early stages are required to establish strong and accurately tailored treatment rationales for frail patients. We analysed participants aged 60 and above using cross-sectional biochemical and survey data from the Berlin Aging Study II (BASE-II, N = 1512, pre-frail=470, frail=14), applying machine-learning techniques to investigate determinants of physical frailty measured by Fried et al.'s 5-item frailty phenotype. Our findings highlight new prognostic sex-specific biomarkers of pre-frailty (the early stage of frailty) with possible clinical applications, enriching the current sex-agnostic diagnostic scores with easy monitorable physical and physiological characteristics. Low appendicular lean mass and high fat composition in men, or vitamin D deficiency and high white blood cell counts in women, emerged as strong indicators of the respective pre-frailty profiles. Because the number of fully frail individuals was extremely small (n = 14, <1 %), our findings should be interpreted as reflecting predictors of pre-frailty, not of frailty itself. We conclude that understanding the development of frailty remains a complex challenge, and that sex-specific differences must be considered by clinical geriatricians and researchers.

Abbreviations: BASE-II, Berlin Aging Study II; BMI, body mass index; WHR, waist-to-hip ratio; ALM, appendicular lean mass; EX, exhaustion; GA, gait; GS, grip strength; PA, physical activity; WL, weight loss; PCA, principal component analysis; RUS, random under-sampling; SVM, support vector machine; ROC, receiver operating characteristics; CV, cross-validation; CI, confidence interval; PM, physical measurements; BF, body fluids; CG, cognition; SMOTE, synthetic minority over-sampling technique; LDA, linear discriminant analysis; BMD, bone mineral density; BMC, bone mineral content; HDL, high-density lipoprotein; MANOVA, multivariate analysis of variance; BCD, between-cluster distance; WCD, within-cluster distance; IM, individual medications; ID, individual devices; GM, grouped medications; SV, surveys; CH, chronic morbidity; GD, grouped devices; HPC, high-performance computing; DOR, diagnostics odds ratio; TP, true positive; TN, true negative; PR, precision-recall; FPR, false positive rate.

* Corresponding author.
E-mail addresses: jeff.didier@uni.lu (J. Didier), sebastien.delandtsheer@uni.lu (S. De Landtsheer), maria.pacheco@uni.lu (M.P. Pacheco), ali.kishk@icloud.com (A. Kishk), jochen.schneider@uni.lu (J.G. Schneider), dgoldeck@yahoo.de (D. Goldeck), graham.pawelec@uni-tuebingen.de (G. Pawelec), dominik.spira@charite.de (D. Spira), ilja.demuth@charite.de (I. Demuth), thomas.sauter@uni.lu (T. Sauter).

https://doi.org/10.1016/j.mad.2025.112114
Received 3 June 2025; Received in revised form 16 September 2025; Accepted 18 September 2025
Available online 24 September 2025

1. Introduction

Frailty is a geriatric medical condition highly associated with age and age-related disorders (Lally and Crome, 2007; Clegg et al., 2013; Fulop et al., 2010; Rockwood et al., 1994; Abizanda et al., 2016). It is characterised by an increased vulnerability to health issues such as chronic or inflammatory diseases (Fulop et al., 2010; Holliday, 1995; Fedarko, 2011) and consequently leading to substantially higher institutionalisation rate. Two systematic reviews of 21 studies with 61,500 community-dwelling individuals (Collard et al., 2012), and 46 observational studies with 100,313 participants (Ofori-Asenso et al., 2019) aged 65 or older, found weighted frailty prevalences of 10.7 % and 13.6 %, respectively, and frail individuals survived a median of 3 years. Frailty impacts over 10 million Europeans and 75 million individuals worldwide, severely deteriorating their quality of life. The European Commission and the United Nations forecasted a sharp rise of the number of older people and the percentage share of the population (From 90.5 (20.3 %) to 129.8 (29.4 %) million and from 727 (9.3 %) million to 1.5 (16.0 %) billion people, respectively) in 2050, posing a severe burden on public health care systems if the underlying causes of frailty development are improperly acted upon (Abreu et al., 2019).

Since the 1990s, several attempts have sought to establish the concept of the frailty phenotype and its pathophysiological features, but did not reach consensus in the medical community (Morley et al., 2013), instead resulting in various theories describing the condition as a stand-alone phenotype (Clegg et al., 2013; Rockwood et al., 1994; Fried et al., 2001), a medical syndrome (Chen et al., 2014; Xue, 2011), or the consequence of deficit accumulations over the lifetime (Rockwood and Mitnitski, 2007). The multifaceted and diverse nature of the putative underlying causes contributing to frailty has led most studies to focus on risk factors and on categorization strategies based on distinct manifestations, comprising physical (Fried et al., 2021; Liu et al., 2017), cognitive aspects (Idris and Badruddin, 2021; Panza et al., 2015), affected organs or systems (Fried et al., 2021; Di Sabatino et al., 2018; Wang et al., 2019; Dibello et al., 2021; Buchmann et al., 2019; Laur et al., 2017; Vaz Fragoso et al., 2012), including but not limited to the liver, gut, oral cavity, metabolism, renal and respiratory systems.

The various suggested subtypes of frailty all share the same fate of irregular and disproportionate decline in physiological reserve compared to healthy ageing (see Fig. 1, inspired from *Patel et al (*Patel et al., 2017*).*), preceded by risk factors such as age, disabilities, co-morbidities, inflammation, falls, poor physical activity, metabolic or cardiovascular disorders, amongst others (see references (Fulop et al., 2010; Abreu et al., 2019; Buchmann et al., 2019; Vaz Fragoso et al.,

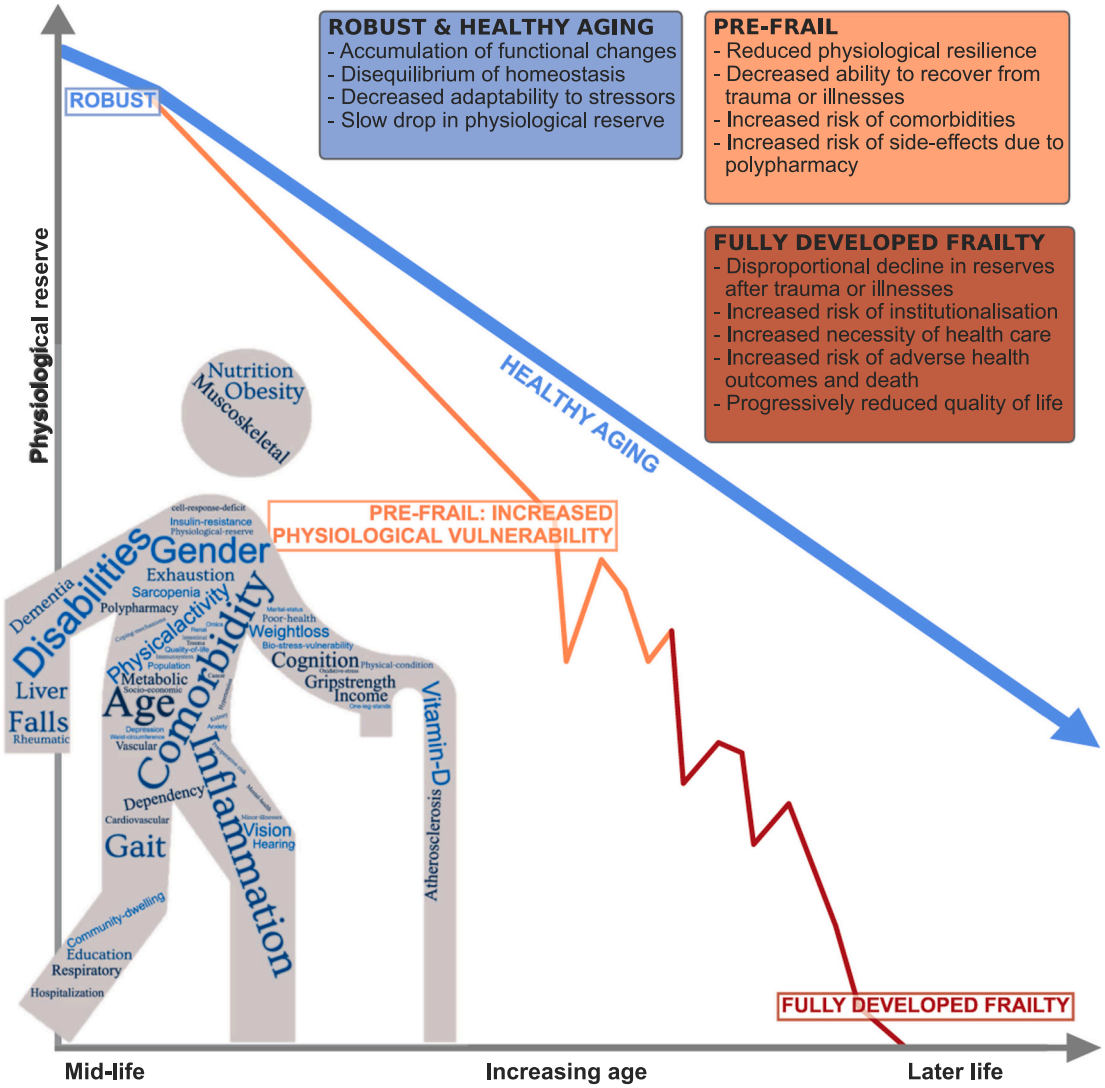


Fig. 1. Healthy ageing versus frailty development. Healthy ageing (blue) and frailty development (orange and red) in terms of physiological reserve; showing the hallmark consequences of the three phases that compose frailty development (robust ageing, pre-frail, and fully developed frailty); and including word cloud with the most prominent frailty risk factors scaled by the appearance in literature. The figure was inspired and adapted from Patel et al. (Patel et al., 2017).

2012; Espinoza et al., 2018; Puts et al., 2005; Spira et al., 2015; Fhon et al., 2018), and the word cloud in Fig. 1). However, measurement of physiological reserve is often limited to physical characteristics such as grip strength, various parameters of the limbs or other body parts, and cardiovascular or cardiopulmonary exercise (Linn et al., 2021; O'Caoimh et al., 2021), neglecting physiological aspects such as age- or muscle-related molecular blood biomarker profiles. Hence, all frailty studies exhibit different strengths and weaknesses, according to a 2021 review by *Howlett et al (*Howlett et al., 2021*)*.. Thus, in practice, the therapies and treatments used to care for frail patients target the symptoms and aim to alleviate the impact on quality of life rather than addressing the underlying causes of frailty (Lally and Crome, 2007). Furthermore, frailty-associated phenotypes including vitamin D deficiency (Shardell et al., 2009; Zhou et al., 2016; Spira et al., 2019; Murad et al., 2011), polypharmacy (Gutiérrez-Valencia et al., 2018; Nwadiugwu, 2020), comorbidities (Espinoza et al., 2018; Tenison and Henderson, 2020), sarcopenia (Patel et al., 2017; Spira et al., 2015), diabetes (Strain et al., 2021), and chronic inflammation (Puts et al., 2005), among others, that are often linked to detrimental effects on the physiological reserve and the individual's ability to recover from insults, are rarely considered by the existing frailty assessment instruments. In this study, we focus on a dataset of older people aged 60 years and above from the observational Berlin Aging Study II (BASE-II) (Bertram et al., 2014). Frailty in BASE-II was defined according to *Fried et al.*'s 5-item frailty phenotype (Fried et al., 2001) (Tab. S1) and shown to be associated with low lean mass, physical performance, sex-specific vitamin D deficiency, metabolic syndrome, use of potentially inappropriate medication, and mortality (Buchmann et al., 2019; Spira et al., 2015, 2019; Toepfer et al., 2021; Koenig et al., 2024). Here, we applied machine-learning approaches to identify prognostic biomedical and clinical markers that can best reflect *Fried et al.*'s frailty phenotype in BASE-II.

Previous studies have harnessed the power of machine-learning algorithms to identify biomarkers and risk factors of frailty development in older people. *Gomez-Cabrero et al.* identified protective biomarkers as well as a risk factor for frailty in a nested case-control study in four European ageing cohorts (Gomez-Cabrero et al., 2021). The protective factors included, among others, vitamin D and lutein zeaxanthin (carotenoid, and precursor of vitamin A). Troponin T, a protein in the heart muscle which is responsible for binding calcium was found to be main risk factor. *Park et al.* investigated the least number of digital biomarkers of frailty measurable with sensors during daily activities (Park et al., 2021). *Akbari et al.* demonstrated that machine-learning modelling can quantify the risk of frailty in older people based on real-time skeletal movements using the Microsoft Kinect-Based Skeleton Pose, where participants performed different physical exercises (sit-and-stand, arm curls, …) in front of a Kinect sensor (Akbari et al., 2021). In 2023, *Da Cunha Leme et al.* applied machine-learning on the English Longitudinal Study of Aging to identify the best predictors of frailty in middle-aged and older adults based on features such as age, balance, household wealth, alcohol intake, and depression (Leme et al., 2023).

In this study, we exploited the rich biomedical and clinical information available in BASE-II to narrow down the most important prognostic biomarkers associated with pre-frailty, thereby validating and supplementing the existing body of knowledge with new insights. Our analyses focused primarily on pre-frail individuals (the initial stage on a trajectory that may eventually lead to fully developed frailty), while recognizing that pre-frail and frail represent two distinct clinical stages. With the impending surge in the older population, added to the persisting lack of a gold standard and consensus on frailty, we examined biomedical and clinical parameters ranging from blood-based biomarkers to disease status and functionality. This approach acknowledges the development of frailty as a systemic condition that manifests across domains, necessitating diverse input variables to identify robust and clinically relevant biomarkers. In addition, this study explored factors relevant to sex-specific pre-frailty trajectories, an under-investigated subject as sex-based biological differences, including hormonal regulation, inflammatory responses, and body composition, may influence frailty development and manifestation. As such, acknowledging sex-specific patterns of pre-frailty patterns could improve risk stratification, early detection, and the development of more effective, personalized interventions.

2. Results

The analyses presented here focus on pre-frail participants, that is, individuals at the initial stage of a trajectory that may eventually lead to fully developed frailty. Throughout the text and figures, the label 'pre-frail/frail' refers overwhelmingly to pre-frail participants (n = 470), with only a very small subset of frail participants (n = 14, <1 %). Consequently, all findings should be interpreted as reflecting predictors of pre-frailty, no conclusions can be drawn about fully developed frailty from this data.

2.1. Descriptive and inferential analysis

2.1.1. Sex-specific differences in the BASE-II description of general frailty-related phenotypes

The BASE-II cohort description of frailty-related phenotypes such as sex, age, body mass index (BMI), waist-to-hip ratio (WHR), appendicular lean mass (ALM), sarcopenia, morbidity index, and polypharmacy revealed interesting differences between non-frail and pre-frail/frail participants when compared across sex (Table 1). Sex itself was not significant in the mixed-sex group (p-value = 5.23e-01), but sharp contrasts emerged, for example:

- WHR (significant in men only, p-value = 9.48e-03)
- ALM (significant in men and mixed-sex, p-values = 3.04e-03 and 2.84e-02, respectively)
- Morbidity index (significant in women and mixed-sex, p-values = 1.43e-04 and 1.72e-04, respectively)
- Polypharmacy (significant in women and mixed-sex, p-values = 1.43e-03 and 3.39e-04, respectively)

Detailed visualizations of the differences in general frailty-related characteristics and their corresponding distributions are depicted in the supplementary section (see Fig. S1 and Fig. S2 for frailty-related features based on data type). As expected, BMI-adjusted ALM, height, ALM, and WHR showed clear sex differences, with men having higher values.

2.1.2. Sex-specificity within Fried et al.'s Frailty phenotype items

Sex-specificity was further investigated across *Fried et al.*'s Frailty Score, frailty phenotype, as well as the five underlying variables of exhaustion (EX), gait (GA), grip strength (GS), physical activity (PA), and weight loss (WL) (Table 2). No statistical significance could be observed in both the Frailty Score (p-value = 7.03e-01) and the frailty phenotype (p-value = 5.23e-01), but significant differences manifested within four of the five *Fried et al.*'s Frailty items, except WL (p-value = 5.61e-01). While reported EX (p-value = 2.94e-03) and weak GS (p-value = 5.77e-04) were more frequent in frail women, GA impairment (p-value = 1.34e-02) and lack of PA (p-value = 3.88e-02) were more frequent in frail men. The clinical variable of GS in particular shows the largest difference between the two sexes, with a corrected Cramér's V coefficient of 0.09 (small correlation considering one degree of freedom), and the lowest p-value of 5.77e-04. These observations suggest that *Fried et al.*'s frailty phenotype manifests differently in men and women, showing that sex-specific interpretation is strongly recommended in the BASE-II frailty analysis.

Table 1
BASE-II cohort characteristics. Context-specific BASE-II cohort characteristics in relation to the frailty phenotype, with the first triplet of columns displaying mixed-sex participants, second men, and third women. P-values obtained with the Cramér's V correlation algorithm corrected for binary and discrete data types. Continuous features underwent Welch's unequal variance T-test after assessing the equality of variances between the non-frail and pre-frail/frail group with the Levene's test. SD standard deviation, BMI body mass index, ALM appendicular lean mass, WHR waist-hip ratio, IQR inter quartile range, ns non-significant, 0.01 < * < 0.05, 0.001 < ** < 0.01, 0.0001 < *** < 0.001, 0.00001 < **** < 0.0001.

Cohort characteristics	Non-frail mixed (N = 1028, 67.99 %)	Pre-frail / frail mixed (N = 484, 32.01 %)	P-value	Non-frail men (N = 513, 33.93 %)	Pre-frail / frail men (N = 233, 15.41 %)	P-value	Non-frail women (N = 515, 34.06 %)	Pre-frail / frail women (N = 251, 16.60 %)	P-value
Sex	513 (33.93 %)	233 (15.41 %)	5.23e−01	/	/	/	/	/	/
men N (%)	515 (34.06 %)	251 (16.60 %)	(ns)						
women N (%)									
Age (years)	68.46 ± 3.44	69.31 ± 4.16	1.05e−04	68.68 ± 3.46	69.76 ± 4.33	8.62e−04	68.25 ± 3.41	68.89 ± 3.96	2.77e−02
mean ± SD	[60.16, 83.25]	[60.74, 84.63]	(***)	[60.16, 78.12]	[60.74, 82.79]	(***)	[61.29, 83.25]	[61.46, 84.63]	(*)
[min, max]									
BMI (kg/m²)	26.37 ± 3.79	27.73 ± 4.72	4.73e−08	26.93 ± 3.45	27.97 ± 3.77	2.13e−04	25.82 ± 4.03	27.50 ± 5.45	1.99e−05
	[17.03, 44.24]	[17.94, 47.68]	(****)	[19.13, 44.24]	[19.97, 42.61]	(***)	[17.03, 40.4]	[17.94, 47.68]	(****)
Waist/Hip ratio	0.96 ± 0.08	0.96 ± 0.08	3.35e−01	1.01 ± 0.05	1.02 ± 0.05	9.36e−03	0.90 ± 0.07	0.91 ± 0.06	6.88e−01
	[0.57, 1.44]	[0.74, 1.21]	(ns)	[0.81, 1.25]	[0.88, 1.17]	(**)	[0.57, 1.44]	[0.74, 1.21]	(ns)
Height (cm)	169.74 ± 8.81	167.84 ± 8.94	1.07e−04	176.09 ± 6.17	174.53 ± 6.21	1.46e−03	163.41 ± 6.06	161.63 ± 6.18	1.68e−04
	[146.0, 194.8]	[144.0, 191.0]	(***)	[160.0, 194.8]	[156.3, 191.0]	(**)	[146.0, 182.5]	[144.0, 176.0]	(***)
ALM (kg)	21.38 ± 4.95	20.80 ± 4.77	3.05e−02	25.49 ± 3.06	24.75 ± 3.17	2.62e−03	17.29 ± 2.46	17.13 ± 2.56	3.99e−01
	[9.28, 35.1]	[11.97, 35.82]	(*)	[18.91, 35.1]	[17.42, 35.82]	(**)	[9.28, 25.31]	[11.97, 26.18]	(ns)
ALM/BMI ratio	0.82 ± 0.18	0.76 ± 0.17	1.54e−08	0.96 ± 0.13	0.89 ± 0.11	3.65e−10	0.68 ± 0.10	0.64 ± 0.11	3.19e−07
	[0.42, 1.47]	[0.38, 1.31]	(****)	[0.63, 1.47]	[0.59, 1.31]	(****)	[0.42, 1.03]	[0.38, 0.93]	(****)
Heart	863	350	1.16e−07	451	185	2.37e−03	412	165	1.73e−05
insufficiency	165	134	(****)	62	48	(**)	103	86	(****)
unaffected	0 (0.0, 0.0)	0 (0.0, 1.0)		0 (0.0, 0.0)	0 (0.0, 0.0)		0 (0.0, 0.0)	0 (0.0, 1.0)	
affected									
median (IQR)									
Vitamin D	537	194	1.02e−05	270	97	5.35e−03	267	97	5.96e−04
deficiency	491	290	(****)	243	136	(**)	248	154	(***)
	0 (0.0, 1.0)	1 (0.0, 1.0)		0 (0.0, 1.0)	1 (0.0, 1.0)		0 (0.0, 1.0)	1 (0.0, 1.0)	
Sarcopenia	968	416	8.69e−08	468	200	2.57e−02	500	216	6.57e−09
	60	68	(****)	45	33	(*)	15	35	(****)
	0 (0.0, 0.0)	0 (0.0, 0.0)		0 (0.0, 0.0)	0 (0.0, 0.0)		0 (0.0, 0.0)	0 (0.0, 0.0)	
Morbidity Index	427	168	1.72e−04	205	82	2.11e−01	222	86	1.98e−04
unaffected	295	128	(***)	133	62	(ns)	162	66	(***)
only one	294	168		169	82		125	86	
two - four	12	20		6	7		6	13	
five or more	1 (0.0, 2.0)	1 (0.0, 2.0)		1 (0.0, 2.0)	1 (0.0, 2.0)		1 (0.0, 2.0)	1 (0.0, 2.0)	
median (IQR)									
Polypharmacy	203	64	3.39e−04	116	35	6.72e−02	87	29	1.43e−03
	157	56	(***)	73	31	(ns)	84	25	(**)
	460	230		216	107		244	123	
	208	134		108	60		100	74	
	2 (1.0, 4.0)	3 (2.0, 5.0)		2 (1.0, 4.0)	2 (1.0, 5.0)		2 (1.0, 4.0)	3 (2.0, 5.0)	

2.1.3. *Sex differences within the BASE-II frailty profiles*

Both sexes were balanced in the non-frail and the pre-frail/frail group, with approximately twice as many non-frail participants (Fig. 2a). Most participants (97.1 %) were affected by one or two of the five *Fried et al.* frailty items, and only 14 participants (3.9 %) affected by three or more items (Fig. 2b). The distribution of the different phenotypes showed that GA and PA were predominantly impaired in men, while EX and GS were predominantly impaired in women (Fig. 2c). Only WL was observed to be balanced, although with the overall lowest count (37 total cases of WL compared to the 115, 133, 137, and 168 cases of GS, PA, EX, and GA, respectively). While most participants (81.2 %) were affected by one single frailty item (Fig. 2d), the intersections revealed that more men are suffering from single GA or PA impairment and combinations thereof, and that single EX or GS affection and combinations thereof are more often observed in women. Although these sex differences in the BASE-II frailty profiles show solid evidence for considering sex-specific modelling of frailty in this cohort, we further analysed the available biomedical data for similar signs using dimensionality reduction. In Fig. 2e, Principal component analysis (PCA) of the continuous data is resulting in two clusters separating both sexes, showing that these differences not only reside within *Fried et al.*'s 5-item frailty phenotype, but are also noticeable across the large panel of biomedical information of BASE-II. Furthermore, since frailty is highly associated with age and age-related disorders, we analysed the age differences in non-frail and pre-frail/frail sexes (Fig. 2f), showing a lower average age in non-frail than pre-frail/frail participants, and specifically lower in women (non-frail men 68.68 ± 3.46 years, pre-frail/frail men 69.76 ± 4.33 years (p-value = 8.62e-04); non-frail women 68.25 ± 3.41 years, pre-frail/frail women 68.89 ± 3.96 years (p-value = 2.77e-02)).

2.2. *Pre-frailty biomarker analysis in BASE-II*

2.2.1. *Complete data-driven model architecture and performances*

Using all the available biomedical information as input (indicated by the affix *ALL*), pre-processing reduced the number of input features to 257 (mixed), 167 (men), and 33 (women) features (see Section 4.4.3 for more details about the low number of input features for the women model, and the summary of feature correlation analysis in Fig. S3a). Among the possible combinations of pipeline architecture and machine-

Table 2
BASE-II cohort characteristics of the 5 *Fried et al.* frailty phenotype items. *Fried et al.* frailty score, frailty phenotype, and frailty items characteristics in the mixed-sex, men, and women BASE-II groups. P values obtained with the Cramér's V correlation algorithm corrected for binary and discrete data types. ns non-significant, $0.01 < * < 0.05$, $0.001 < ** < 0.01$, $0.0001 < *** < 0.001$, $0.00001 < **** < 0.0001$.

BASE-II Fried et al. frailty phenotype description	Mixed (N = 1512, 100 %)	Men (N = 746, 49.34 %)	Women (N = 766, 50.66 %)	P-value
Fried et al. Frailty Score	1028	513	515	7.03e−01
unaffected	393	188	205	(ns)
one	77	39	38	
two	13	51	80	
three	1			
four				
Fried et al. frailty phenotype	1028	513	515	5.23e−01
unaffected	484	233	251	(ns)
affected				
Exhaustion	1375	695	680	2.94e−03
(poor endurance)	137	51	86	(**)
Gait	1344	648	696	1.34e−02
(slowness)	168	98	70	(*)
Grip strength	1397	707	690	5.77e−04
(weakness)	115	39	76	(***)
Physical activity	1379	669	710	3.88e−02
(low activity)	133	77	56	(*)
Weight loss	1475	726	749	5.61e−01
(shrinking)	37	20	17	(ns)

learning classifier parameters (detailed in Tab. S3 point 3 and Tab. S4; see Fig. 3e for a summary of the best configurations and parameters), the best performing *ALL*-models each share the use of random under-sampling (RUS) technique to balance pre-frail/frail and non-frail participants, PCA to reduce the input dimensionality, and χ^2 to select the top discrete features. They differ however by the technique of scaling the continuous input (Min-max scaling for mixed-sex and men, standard scaling for women) and the type of support vector machines (SVM) kernel applied by the classifier (radial basis function for *mixed-ALL*, linear for *men-ALL* and *women-ALL*). Regularization values were 3000 for *mixed-ALL*, 0.1 for *women-ALL*, and 100 for the best *men-ALL* model. The classifier tolerance was the highest for *men-ALL* with 0.01, lowest for *mixed-ALL* with 0.00001, and 0.001 for the best performing *women-ALL* model. In Fig. 4a, the performances are depicted by the receiver operating characteristics (ROC) curves of the 5-times 10-fold cross-validation (CV), the 20 % hold-out test set, and the corresponding hold-out confusion matrix for all three models. The best *mixed-ALL* model reached a CV ROC mean of 0.665 (95 % confidence interval (CI): [0.551 – 0.779]) with a hold-out ROC score of 0.63. In terms of the hold-out confusion matrix, 72.73 % non-frail and 53.85 % pre-frail/frail participants were correctly identified. This mixed-sex model was moderately outperformed by both the *men-ALL* and *women-ALL* models, given their CV ROC mean of 0.705 (95 % CI: [0.567 – 0.843]) and 0.693 (95 % CI: [0.551 – 0.835]), respectively, and both yielded the same hold-out test ROC score of 0.66. The *men-ALL* model correctly identified 64.29 % non-frail and 67.27 % pre-frail/frail participants, whereas the *women-ALL* model identified 69.00 % non-frail and 63.27 % pre-frail/frail participants correctly. Given that sex-specific models outperformed the mixed-sex models, we performed subgroup analysis to gain potentially deeper insights into sex-specific pre-frailty.

2.2.2. Subgroup-based models outperform complete models for pre-frailty prediction in men and women

In a second experiment, the interest was focused on four of the ten biomedical subgroups defined in Section 4.2, namely nutrients (NT), physical measurements (PM), body fluids (BF), and cognition (CG), as highlighted in Fig. 3a. The best performing subgroup for each context was recorded and will be hereafter indicated by their respective affix as *mixed-BF, men-PM*, and *women-BF*, which were reduced to 86, 32, and 84 input features, respectively, after pre-processing. (refer to Section 4.4.3 for more details about the low number of input features for the men model, and the summary of feature correlation analysis in Fig. S3b). Like the complete data-driven models, the subgroup models share some specific characteristics while strongly differing in others. These models' configuration commonly applies the same strategy for balancing out pre-frail/frail participants; however, this time synthetic minority over-sampling technique (SMOTE) was prioritized. Moreover, all three subgroup models performed best using the linear SVM kernel after dimensionality reduction. The *mixed-BF* and *women-BF* models used robust scaling of the continuous input data, PCA to reduce the input dimensionality, and χ^2 for selecting the top discrete features from the input. The *men-PM* model used standard scaling of the input features and linear discriminant analysis (LDA) to reduce the input dimensionality to one single feature distinguishing non-frail from pre-frail/frail. In Fig. 4b, the performance of the best mixed-sex model (*mixed-BF*) reached a CV ROC mean of 0.629 ± 0.047 (95 % CI: [0.535 – 0.723]) with a hold-out ROC score of 0.61. 61.62 % non-frail and 59.62 % pre-frail/frail participants were correctly identified. This best performing *mixed-BF* model was significantly outperformed by the *men-PM* and moderately outperformed by the *women-BF* model given their CV ROC mean of 0.721 ± 0.066 (95 % CI: [0.619 – 0.853]) and 0.633 ± 0.068 (95 % CI: [0.497 – 0.769]), respectively. The *men-PM* model yielded the highest hold-out test ROC score of 0.70 while the *women-BF* model scored 0.67. The *men-PM* model correctly identified 68.37 % non-frail and 70.91 % pre-frail/frail participants whereas the *women-BF* model identified 63.00 % non-frail and 71.43 % pre-frail/frail participants correctly. With the improvements observed in the women and especially men models based on distinct biomedical subgroups and model architecture, the top contributing features were interpreted context-wise to investigate their underlying associations with the sex-specific pre-frailty profiles observed in BASE-II.

2.2.3. Mixed-sex and sex-specific model interpretability

From the best performing *ALL*-models, the 30 most-contributing features based on their permutation score are depicted in Fig. 4c ($n_{permutation}$ mixed-sex = 1000, men = 500, women = 500). A more detailed view on the feature permutation behaviour in the complete data-driven models can be found in Fig. S4 of the supplementary section. Furthermore, red and green squares shown next to the first ten features in Fig. 4c indicate the direction of the association (green: positive, red: negative), useful to interpret the model's behaviour and decision from a biological perspective. Fig. 4c show the performance of the subgroup-based data-driven models using *PM*-, *BF*-, *NT*-, and *CG*-related input features, with values displayed only for the complete and the best performing subgroup models. The most contributing features of the subgroup data-driven models are depicted in Fig. 4d, with the corresponding feature permutation behaviour in Fig. S5. Besides the feature importance and permutation behaviour, an additional layer of model interpretability is given by revealing the predictions in terms of the true *Fried et al.* frailty levels (non-frail, pre-frail, and frail) instead of the combined pre-frail and frail, showing that the models performances do increase with the level of true frailty (Fig. 6). As a result, the best *mixed-BF* model correctly identified pre-frail and frail participants with sensitivities of 60.0 % and 85.7 % respectively. In *men-PM*, the pre-frail and frail sensitivities are 70.5 % and 83.3 % respectively, while the best *women-BF* model predicted pre-frail and frail participants with sensitivities of 63.8 % and 87.5 % respectively. The false positive rates for the non-frail participants of those three top models are 41.4 %, 35.5 %, and 41.9 %, respectively, with specificities of 58.9 % in *mixed-BF*, 64.5 % in *men-PM*, and 58.1 % in *women-BF*. This stratified analysis shows that the identified predictors remain significant in the small frail subgroup.

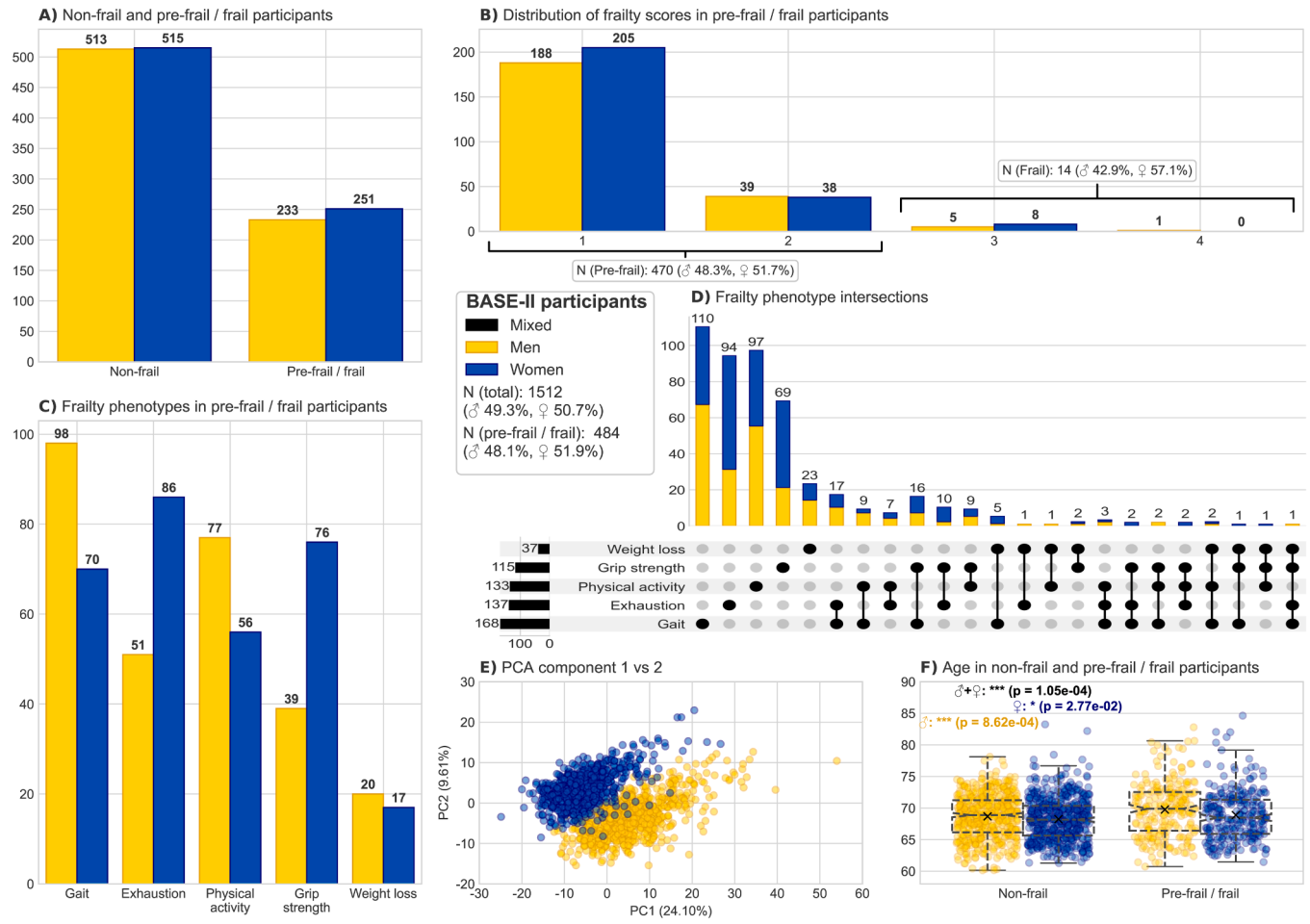


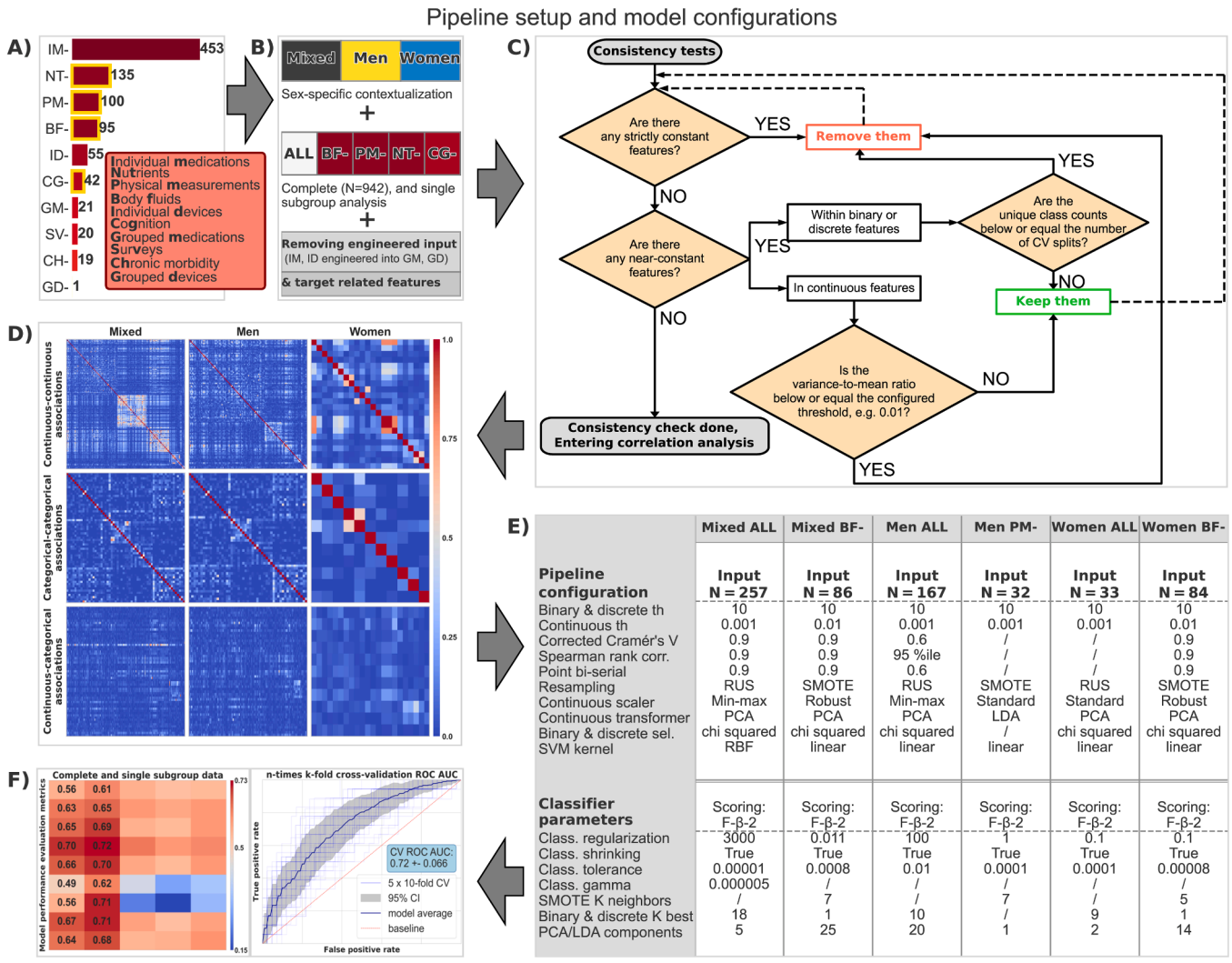
Fig. 2. Sex-specific differences in the BASE-II participants' pre-frailty profile call for contextualized modelling. Distribution of the non-frail and pre-frail/frail participants (A), the *Fried et al.* Frailty Score equal or above 1 (B); the *Fried et al.* frailty phenotype items (C) and phenotype intersections (D) show sex-specific frailty profiles. The first and second principal components (E) clearly separate both sex; and the age in non-frail and pre-frail/frail participants do also show significant differences (F). Statistical significance derived from Welch's unequal variance T-test. Showing mixed-sex (black), men (yellow), and women (blue) participants of BASE-II. PCA principal component analysis, $0.01 < * < 0.05$, $0.001 < ** < 0.01$, $0.0001 < *** < 0.001$, $0.00001 < **** < 0.0001$.

2.2.4. Sex-specific pre-frailty patterns share morbidities but diverge in cognitive and physical traits

The three most important features driving the *mixed-ALL* model are heart insufficiency, vitamin D deficiency, and sarcopenia, three independent morbidities and deficits linked to cardiovascular, physiologic, and muscular pathologies, respectively. The red marking of those three features in Fig. 4c suggests that the more deficits the participants (men and women alike) have accumulated, the more likely they are to be pre-frail/frail (Fig. S6, training set only). Furthermore, the performance in cognitive tasks, particularly the digit symbol substitution test, was weaker in pre-frail/frail participants. The next most contributing features in the *mixed-ALL* context include multiple body fat composition characteristics, particularly of the head (fat percentage), left arm (fat mass in grams), left leg (fat percentage and mass in grams), and the ALM-to-BMI ratio. Pre-frail/frail participants show higher fat composition in the above-mentioned body parts and lower ALM-to-BMI ratio. The top ten contributing features also included docosanoic (behenic) acid, a non-essential long-chain saturated fatty acid, which appeared to be consumed less by pre-frail/frail participants and may be associated with increased cardiovascular mortality (Tao et al., 2024).

Compared to the mixed-sex model, the top ten contributing features in *men-ALL* include affection by heart insufficiency and larger head fat mass in pre-frail/frail men. Again, heart insufficiency is the most contributing feature and moreover shows a stronger detrimental effect to the model performance after feature permutation as in the mixed-sex model (comparing their loss in CV score). Although this is also observed for the head fat mass feature, its low z-score indicates lower reliability in this effect. The *men-ALL* model highlights two medication types targeting distinct disease areas: gastroenterology and ophthalmology. Intriguingly, the model behaviour reveals that there is no difference between the non-frail and pre-frail/frail men in terms of the number of medications targeting these two disease areas, and the proportion of pre-frailty is not increasing with the number of medications, even when considering their accumulation (Fig. S7, training set only). This is different in women, as here the proportion of pre-frail participants increases with the number of ophthalmology-related medications and the accumulation with gastroenterology-related medications.

Regarding the *women-ALL* model, two of the top ten contributing features are shared with the *mixed-ALL* model, namely heart insufficiency and sarcopenia. Another indication of the latter deficit is the ALM, which was picked up by the *women-ALL* model, but not in the *mixed-ALL* or *men-ALL* model. Marked as green, the model shows that pre-frail/frail women appear to have lower ALM, or, in other words, lower muscle mass in limbs compared to non-frail women. On the other hand, the *women-ALL* model also shares two common features with the *men-ALL* model, namely again heart insufficiency (which is the only common feature across all three contexts), and the increased time necessary to solve the cognitive trail making test B. The *women-ALL* model is further driven by various drinking habits (increased number of distilled drinks and self-reported drinking frequency), medication intake



	Mixed ALL	Mixed BF-	Men ALL	Men PM-	Women ALL	Women BF-
Pipeline configuration	**Input N = 257**	**Input N = 86**	**Input N = 167**	**Input N = 32**	**Input N = 33**	**Input N = 84**
Binary & discrete th	10	10	10	10	10	10
Continuous th	0.001	0.01	0.001	0.001	0.001	0.01
Corrected Cramér's V	0.9	0.9	0.6	/	/	0.9
Spearman rank corr.	0.9	0.9	95 %ile	/	/	0.9
Point bi-serial	0.9	0.9	0.6	/	/	0.9
Resampling	RUS	SMOTE	RUS	SMOTE	RUS	SMOTE
Continuous scaler	Min-max	Robust	Min-max	Standard	Standard	Robust
Continuous transformer	PCA	PCA	PCA	LDA	PCA	PCA
Binary & discrete sel.	chi squared	chi squared	chi squared	/	chi squared	chi squared
SVM kernel	RBF	linear	linear	linear	linear	linear
Classifier parameters	Scoring: F-β-2	Scoring: F-β-2	Scoring: F-β-2	Scoring: F-β-2	Scoring: F-β-2	Scoring: F-β-2
Class. regularization	3000	0.011	100	1	0.1	0.1
Class. shrinking	True	True	True	True	True	True
Class. tolerance	0.00001	0.0008	0.01	0.0001	0.0001	0.00008
Class. gamma	0.000005	/	/	/	/	/
SMOTE K neighbors	/	7	/	7	/	5
Binary & discrete K best	18	1	10	/	9	1
PCA/LDA components	5	25	20	1	2	14

Fig. 3. Pipeline setup and context-specific model configurations. The vast amount of biomedical data in BASE-II was divided in 10 subgroups that represent the nature of the biomedical information (A). In a first step, the data is partitioned by sex (mixed-sex, men, women) and either complete or 4 single subgroup analysis (highlighted in A), including the removal of input features that have been used to engineer new ones as well as target related features (B). Next, the filtered data will undergo a consistency test to check for constant and near-constant features according to the configurations (C). The remaining consistent data then undergoes data type-specific correlation analysis to identify and remove mutually correlated features to avoid redundant information according to user-defined parameters (D). The pre-processed data then enters the machine-learning-based classifier pipeline with the predefined configurations and parameter ranges, while showing here only the parameters of the best performing context-specific models (E). Finally, the best performing model selected by exhaustive grid search is evaluated using multiple metrics, and feature importance are measured as loss or gain in CV score during n-times feature permutation (F). IM individual medications, NT nutrients, PM physical measurements, BF body fluids, ID individual devices, CG cognition, GM grouped medications, SV surveys, CH chronic morbidity, GD grouped devices, th threshold, corr correlation, sel selection, SVM Support Vector Machine, SMOTE synthetic minority oversampling technique, PCA principal component analysis, LDA linear discriminant analysis, RUS random under-sampling, RBF radial basis function, CV cross-validation, ROC receiver operating characteristics, DOR diagnostics odds ratio, TP true positive, TN true negative.

targeting dental issues, and decreased cognitive test scores such as the correct wordlist learning total score and the wordlist discriminability z-score. While above observations are revealing first subtle distinctions between mixed-sex, men, and women frailty trajectories in terms of context-specific top contributing features and the interpretation thereof, the subgroup-based data-driven models can offer further insights into the underlying causes and consequences of sex-specific pre- frailty as observed in BASE-II.

2.2.5. *Subgroup-based data-driven models highlight sex-specific physiological and physical traits*

Focusing on the subgroup-based data-driven models, the results hint at intrinsic differences in the pre-frailty profiles in the mixed-sex and sex-specific BASE-II population based on the improved model performance using BF input data for the mixed-sex and women models, and PM for men (Fig. 4c, heatmaps). Compared to the *mixed-ALL* model, the *mixed-BF* model is foremost attuned to vitamin D deficiency, low vitamin B12 and sodium levels, as well as increased immune markers such as circulating IL-10 and monocytes, suggesting an increased burden of underlying medical conditions or complications in pre-frail/frail participants (Fig. 4d).

Substantial improvements in prediction performance were achieved when restricting the men model to PM data only, with nine of the top ten contributing features causing an average loss of CV score of 5–15 % during the permutation importance. Similar observations were made in the *mixed-ALL* scenario, however, showing much more importance in men. Again, multiple body composition features are increased, especially trunk fat, trunk mass, and left arm fat (percentage and grams), while lean mass-related features such as left arm lean mass (in grams), ALM (in grams), and ALM-BMI ratio are decreased in pre-frail/frail

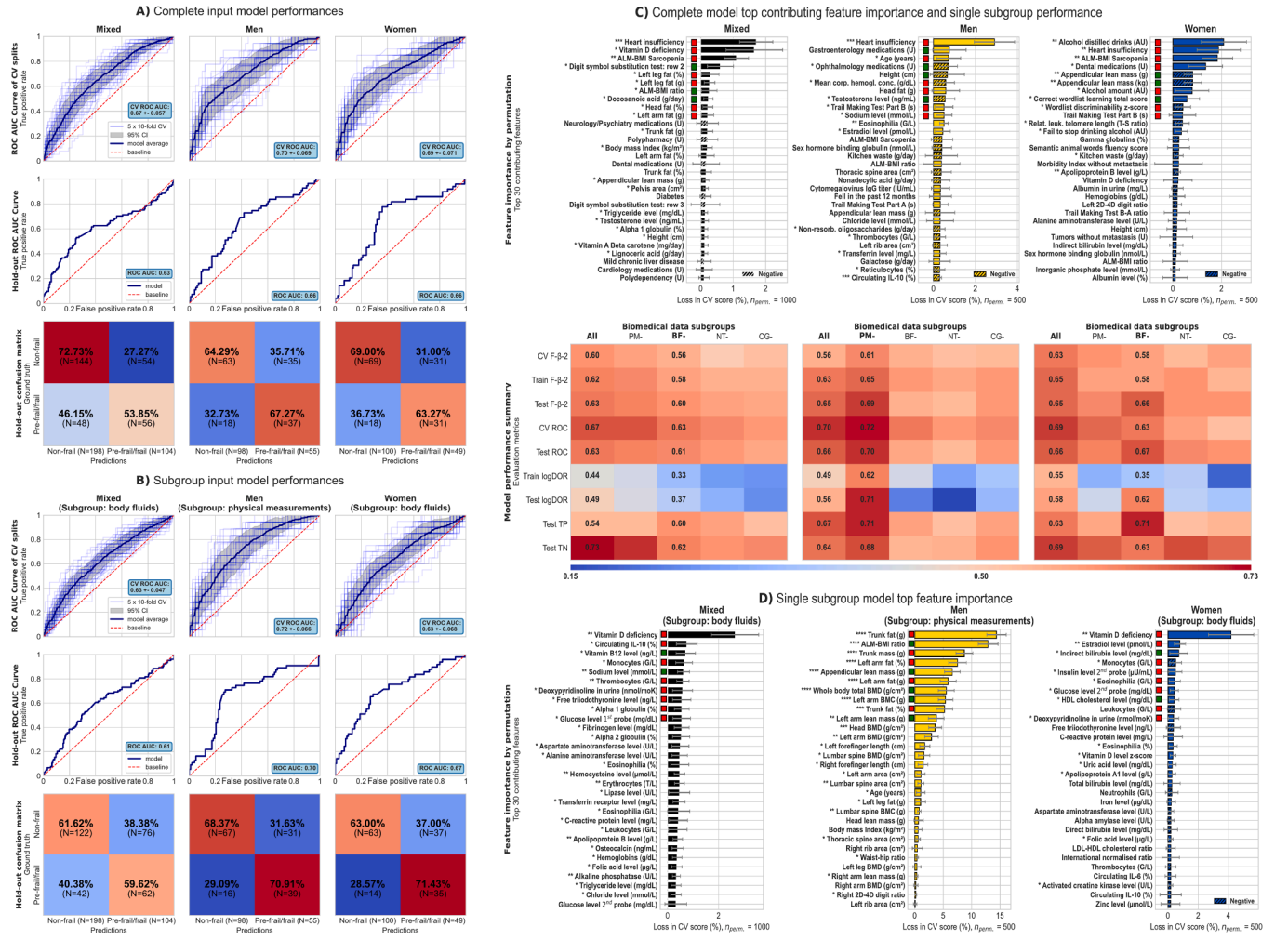


Fig. 4. Sex-specific data-driven models outperform mixed-sex models and reveal distinct predictive patterns for pre-frailty in BASE-II. (A) Context-specific model performances in the complete data-driven models, showing the ROC AUC curves from a 5 times 10-fold CV training process [mixed-sex: 0.665 ± 0.057, men: 0.705 ± 0.069, women: 0.693 ± 0.071] and performance on the hold-out test set [mixed-sex: 0.61, men: 0.65, women: 0.68]. (B) Context-specific model performances in the single subgroup data-driven models, showing ROC AUC curves from the 5 times 10-fold CV training process [mixed-sex: 0.629 ± 0.047, men: 0.721 ± 0.066, women: 0.633 ± 0.068] and on the hold-out test set [mixed-sex: 0.60, men: 0.68, women: 0.71]. The CV splits are not identical to the 10-fold splits used in the pipeline (see Fig. 3e). Final row displays the confusion matrix on the hold-out test set for mixed-sex, men, and women models. (C) Sex-specific most predictive features in the complete data-driven models, based on mean and standard deviation of the n-times permuted feature CV score loss. Negative hatched bars indicate gain in CV score. Green and red squares denote the classifier's interpretation of the feature as lower or higher, respectively, in frail-predicted participants compared to non-frail. Asterisks prefixing feature names represent corresponding absolute z-score levels. (D) Sex-specific most predictive features in the single subgroup data-driven models, following the same conventions as in (C). The second row shows the overall heatmap of evaluation scores for the best-performing models in the complete (All) and single subgroup settings. Final row is displaying the confusion matrix on the hold-out test set. The columns represent the mixed-sex, men, and women groups respectively. ROC receiver operating characteristics, AUC area under the curve, ALM appendicular lean mass, BMI body mass index, T-S telomere-single gene ratio, 2D-4D digit 2-to-digit 4 ratio, ROC receiver operating characteristics, DOR diagnostics odds ratio, TP true positive, TN true negative, BMC bone mineral content, BMD bone mineral density, HDL high density lipoprotein, LDL low density lipoprotein.

participants. In addition, whole-body total bone mineral density (BMD, in grams per square centimetre) and the left arm bone mineral content (BMC, in grams) are decreased in pre-frail/frail men. With this, the pre-frail/frail men profile in BASE-II emerges as a participant with higher fat mass in limbs and trunks, and reduced ALM across the body. Furthermore, their low BMD and BMC can be associated to more fragile bone structures, amplifying the risk of serious medical condition, structural instability, and facilitating bone fractures.

In the *women-BF* model, vitamin D deficiency is the leading driving factor, significantly impacting the CV score during permutation measurement compared to other models. This implies that a deficiency in vitamin D ($\leq$ 50 nmol/L of 25-hydroxy-vitamin D in serum) among women goes along with increased likelihood of developing frailty. Furthermore, multiple white blood cell counts are increased in pre-frail/frail women, i.e., monocytes, leukocytes, and eosinophils (immune functions); but also, indirect bilirubin (clearance function) and high-density lipoprotein (HDL) cholesterol levels (endothelial function) are decreased. These participants show higher levels of oestradiol (hormone and bone health function), insulin (glucose regulation function), and glucose in serum (energy function), as well as higher deoxypyridinoline in urine (bone resorption function). These imbalances portrait a women pre-frailty profile focussing on internal physiological signals whereas pre-frailty in men is more likely revealed by external physical characteristics.

2.2.6. *Sex-specific aggregated feature associations to pre-frailty*

To better understand the association between frailty and specific features in the different models, statistical analysis was conducted on the top ten and thirty contributing features (Table 3a-b and Tab. S5a-b). This revealed that frailty is associated with aggregated, not always

Table 3
BASE-II cohort characteristics of the 10 most predictive features of the complete and single subgroup data-driven models show significant differences in pre-frail/frail and non-frail participants. Cohort characteristics of the top 10 most predictive features in the complete (top 10) and single subgroup (bottom 10) data-driven men (A) and women (B) models respectively. P values obtained with the Cramér's V correlation algorithm corrected for binary and discrete data types. Continuous features underwent Welch's unequal variance T-test after assessing the equality of variances between the non-frail and pre-frail/frail group with the Levene's test. SD standard deviation, BMI body mass index, ALM appendicular lean mass, BMD bone mineral density, BMC bone mineral content, ns non-significant, $0.01 < * < 0.05$, $0.001 < ** < 0.01$, $0.0001 < *** < 0.001$, $0.00001 < **** < 0.0001$.

A) Most important features (men-ALL)	Non-frail men (N = 513, 33.93 %)	Pre-frail / frail men (N = 233, 15.41 %)	P-value	B) Most important features (women-ALL)	Non-frail women (N = 515, 34.06 %)	Pre-frail / frail women (N = 251, 16.60 %)	P-value
Heart insufficiency	0: 451 1: 62	0: 185 1: 48	2.37e−03 (**)	**Alcohol distilled drinks (AU)**	0: 320 1: 148 2: 27 3: 13 4: 5 5: 2	0: 147 1: 73 2: 14 3: 11 4: 4 5: 2	6.46e−01 (ns)
Gastroenterology medications (U)	0: 416 1: 83 2: 14	0: 199 1: 32 2: 2	1.65e−01 (ns)	**Heart insufficiency**	0: 412 1: 103	0: 165 1: 86	1.73e−05 (****)
Age (years) mean ± SD; [min, max]	68.68 ± 3.46 [60.16, 78.12]	69.76 ± 4.33 [60.74, 82.79]	8.62e−04 (***)	**ALM BMI Sarcopenia**	0: 500 1: 15	0: 216 1: 35	6.57e−09 (****)
Ophthalmology medications (U)	0: 465 1: 36 2: 12	0: 214 1: 17 2: 2	3.84e−01 (ns)	**Dental medications (U)**	0: 431 1: 82 2: 2	0: 181 1: 68 2: 2	8.58e−04 (***)
Height (cm) mean ± SD; [min, max]	176.09 ± 6.17 [160.0, 194.8]	174.53 ± 6.21 [156.3, 191.0]	1.46e−03 (**)	**Appendicular lean mass (g)** mean ± SD; [min, max]	17290.69 ± 2457.20 [9284.27, 25307.89]	17128.66 ± 2564.11 [11970.53, 26183.56]	3.99e−01 (ns)
Mean corp. hemogl. conc. (g/dL) mean ± SD; [min, max]	34.43 ± 0.99 [31.3, 37.5]	34.43 ± 1.00 [31.3, 37.6]	9.66e−01 (ns)	**Appendicular lean mass (kg)** mean ± SD; [min, max]	17.29 ± 2.46 [9.28, 25.31]	17.13 ± 2.56 [11.97, 26.18]	3.99e−01 (ns)
Head fat (g) mean ± SD; [min, max]	1171.00 ± 147.21 [861.82, 1710.11]	1189.52 ± 157.57 [826.91, 1663.2]	1.20e−01 (ns)	**Alcohol amount (AU)**	0: 454 1: 58 2: 3	0: 211 1: 35 2: 4	1.53e−01 (ns)
Testosterone level (ng/mL) mean ± SD; [min, max]	4.65 ± 1.97 [0.05, 14.19]	4.40 ± 1.94 [0.05, 11.69]	1.08e−01 (ns)	**Correct wordlist learning total score** mean ± SD; [min, max]	23.18 ± 2.79 [15.0, 30.0]	22.74 ± 3.18 [9.0, 29.0]	5.20e−02 (ns)
Trail Making Test Part B (s) mean ± SD; [min, max]	89.35 ± 34.73 [35.0, 266.0]	95.28 ± 37.92 [28.0, 295.0]	3.63e−02 (*)	**Wordlist discriminability z score** mean ± SD; [min, max]	0.12 ± 0.70 [−3.16, 0.95]	0.11 ± 0.72 [−2.54, 0.87]	8.98e−01 (ns)
Sodium level (mmol/L) mean ± SD; [min, max]	139.35 ± 2.68 [125.0, 147.0]	139.51 ± 2.76 [131.0, 148.0]	4.57e−01 (ns)	**Trail Making Test Part B (s)** mean ± SD; [min, max]	84.84 ± 29.63 [36.0, 238.0]	94.49 ± 33.71 [42.0, 233.0]	5.80e−05 (****)
Most important male features (men-PM)	**Non-frail men** (N = 513, 33.93 %)	**Pre-frail / frail men** (N = 233, 15.41 %)	**P-value**	**Most important features (women-BF)**	**Non-frail women** (N = 515, 34.06 %)	**Pre-frail / frail women** (N = 251, 16.60 %)	**P-value**
Trunk fat (g) mean ± SD; [min, max]	13812.02 ± 4071.45 [3965.92, 28361.59]	15024.44 ± 4742.30 [4878.38, 33258.54]	3.72e−04 (***)	**Vitamin D deficiency**	0: 267 1: 248	0: 97 1: 154	5.96e−04 (***)
ALM BMI ratio mean ± SD; [min, max]	0.96 ± 0.13 [0.63, 1.47]	0.89 ± 0.11 [0.59, 1.31]	3.65e−10 (****)	**Estradiol level (pmol/L)** mean ± SD; [min, max]	35.73 ± 59.27 [9.2, 817.2]	44.03 ± 60.11 [9.2, 471.4]	7.03e−02 (ns)
Trunk mass (g) mean ± SD; [min, max]	41614.36 ± 6726.38 [24496.4, 72459.25]	42853.48 ± 8159.18 [25029.34, 74016.79]	4.34e−02 (*)	**Indirect bilirubin level (mg/dL)** mean ± SD; [min, max]	0.44 ± 0.15 [0.18, 1.22]	0.42 ± 0.15 [0.11, 1.04]	8.40e−02 (ns)
Left arm fat (%) mean ± SD; [min, max]	31.03 ± 5.85 [13.71, 52.38]	33.04 ± 5.82 [16.98, 51.9]	1.53e−05 (****)	**Monocytes (G/L)** mean ± SD; [min, max]	0.40 ± 0.14 [0.16, 1.1]	0.43 ± 0.16 [0.13, 1.51]	2.41e−02 (*)
Appendicular lean mass (g) mean ± SD; [min, max]	25486.42 ± 3064.90 [18910.53, 35100.72]	24747.02 ± 3173.83 [17423.04, 35815.85]	2.62e−03 (**)	**Insulin level 2nd probe (μU/mL)** mean ± SD; [min, max]	59.35 ± 50.15 [5.45, 460.1]	62.76 ± 52.07 [6.86, 502.5]	3.83e−01 (ns)
Left arm fat (g) mean ± SD; [min, max]	1422.53 ± 405.07 [445.24, 3374.5]	1494.40 ± 435.79 [615.73, 3556.75]	2.86e−02 (*)	**Eosinophilia (G/L)** mean ± SD; [min, max]	0.15 ± 0.10 [0.01, 0.9]	0.18 ± 0.15 [0.01, 1.46]	4.30e−03 (**)
Whole body total BMD (g/cm²) mean ± SD; [min, max]	1.24 ± 0.11 [0.94, 1.62]	1.23 ± 0.10 [0.96, 1.65]	1.11e−01 (ns)	**Glucose level 2nd probe (mg/dL)** mean ± SD; [min, max]	109.05 ± 29.00 [38.0, 278.0]	114.24 ± 37.36 [56.0, 333.0]	3.53e−02 (*)
Left arm BMC (g) mean ± SD; [min, max]	206.54 ± 33.64 [115.55, 484.77]	199.43 ± 33.22 [121.59, 314.21]	7.43e−03 (**)	**High density lipoprotein cholesterol level (mg/dL)** mean ± SD; [min, max]	70.18 ± 16.03 [35.0, 134.0]	67.68 ± 16.98 [32.0, 153.0]	4.81e−02 (*)
Trunk fat (%) mean ± SD; [min, max]	32.62 ± 5.57 [14.66, 47.71]	34.39 ± 5.36 [17.46, 51.87]	5.02e−05 (****)	**Leukocytes (G/L)** mean ± SD; [min, max]	5.56 ± 1.44 [2.7, 11.8]	5.92 ± 2.02 [2.1, 24.4]	4.95e−03 (**)

(*continued on next page*)

Table 3 (*continued*)

A) Most important features (men-ALL)	Non-frail men (N = 513, 33.93 %)	Pre-frail / frail men (N = 233, 15.41 %)	P-value	B) Most important features (women-ALL)	Non-frail women (N = 515, 34.06 %)	Pre-frail / frail women (N = 251, 16.60 %)	P-value
Left arm lean mass (g) mean ± SD; [min, max]	3123.96 ± 487.39 [1737.97, 4738.63]	2991.33 ± 504.83 [1911.19, 4546.69]	6.95e−04 (***)	**Deoxypyridinoline in urine (nmol/moK)** mean ± SD; [min, max]	59.34 ± 24.88 [1.0, 191.0]	60.55 ± 25.29 [3.5, 152.0]	5.27e−01 (ns)

individual, features. The top driving sex-specific features (Fig. S9a-b) were also compared within mixed-sex, men, and women using the training dataset only. Some models (e.g., *men-PM*) showed more significant associations compared to the combined training and test set, while other models showed only weaker associations (e.g., *women-BF*). The associations to pre-frailty of the top ten contributing features were then put in perspective with all included features of the BASE-II cohort by comparing the difference in z-score for continuous features (Fig. S10a) and the corrected Cramér's V for binary and categorical features (Fig. S10b) within the training set. Highlighting the top thirty features from the best models (*mixed-ALL, men-PM, women-BF*) revealed that only a fraction of those individual features is significant, suggesting that these features may show aggregated association to pre-frailty rather than individually.

Multivariate analysis of variance (MANOVA) of all combinations of the top ten features in *men-PM* and *women-BF* revealed increasing statistical significance and increasing "1 - Wilk's lambda" score with the number of features tested (Fig. 5a-b), supporting the idea of sex-specific aggregated feature associations to pre-frailty. Similar behaviour was observed in the complete data-driven and the mixed-sex models (Fig. S11). When comparing the between-cluster distance (BCD), within-cluster distance (WCD), and their ratio in the top combinations of raw, PCA, and LDA transformed values, the clearest separation of non-frail and pre-frail/frail men was seen with the LDA transformation of the top combination of five (BCD-WCD ratio: 1.03) to all ten features (BCD-WCD ratio: 0.94); which has also been the feature reduction technique configured for the *men-PM* model (see Fig. 3e). For raw values and PCA-transformed values, the best separation was observed with the top two features ALM-BMI ratio and whole-body total BMD, achieving a BCD-WCD ratio of 0.94 with raw values, and 0.67 with PCA transformed values. Within the *women-BF* model, the differences between the combinations were less clear, yielding the highest BCD-WCD ratio with all ten features transformed by LDA (BCD-WCD ratio: 0.78). However, the best performing feature reduction technique for this model was PCA, which yielded the largest separation using the top combination of five features (BCD-WCD ratio: 0.54). Overall, this analysis suggests that the machine-learning algorithm indeed identified sex-specific aggregations of clinical features that best characterise physical pre-frailty in older men and women.

3. Discussion

3.1. Overview of sex-specific pre-frailty prediction using biomedical profiles

Our results showed that pre-frailty (early stage of frailty), as defined by the *Fried et al.* frailty phenotype, can be characterized by sex-specific factors in people aged ≥ 60 years. These observations support our hypothesis of specific men and women frailty profiles, suggesting sex-specific prevention and treatment approaches. The current literature on frailty often addresses multiple subtypes of frailty and mostly within a mixed-sex population; only in recent years have researchers started to look deeper into frailty by revealing sex-specific interventions (Reid et al., 2022), prevalence (Hessey et al., 2020), as well as sex-specific associations of frailty with mortality (Dallmeier et al., 2020; Verschoor et al., 2024), CG (Karanth, 2023), and socioeconomic status (Wang et al., 2024). However, sex-specific prognostic biomarkers for pre-frailty are still largely unexplored. Furthermore, most machine-learning-based approaches to predict frailty have focussed on movement data, self-reported lifestyle habits, or volatile physical characteristics such as GS or chair-sitting exercises (Park et al., 2021; Akbari et al., 2021; Leme et al., 2023), with only a few diving deeper into the biomedical profile of frailty, and then only in mixed-sex populations (Gomez-Cabrero et al., 2021). Here, we analysed machine-learning-based cross-validated prediction performance in the mixed-sex, men, and women population of BASE-II using biomedical data, as well as specific subgroups based on the origin of the biomedical information to predict the *Fried et al.* frailty phenotype (Fried et al., 2001). While frailty was defined using the *Fried et al.* physical frailty phenotype, the included predictors (e.g., inflammatory markers, chronic diseases, medications, or cognitive performance) span domains more aligned with the *Rockwood* frailty index and accumulation of deficit model (Rockwood et al., 1994; Rockwood and Mitnitski, 2007). This conceptual differences between the outcome definition and the feature space provided additional motivation to go beyond a global model and systematically explore frailty prediction within the distinct biomedical subgroups of BF, PM, CG, or NT. By doing so, we aimed to better understand how different physiological systems contribute to pre-frailty in a sex-specific context, and to identify whether certain domains disproportionately drive predictive performance for men versus women. We therefore position our work as exploratory and hypothesis-generating, with the goal of uncovering interpretable biomedical patterns that can inform future mechanistic or intervention studies.

3.2. Sex-specific differences in the expression of pre-frailty

Descriptive analysis revealed first hints of sex-specificity within this cohort. Although sex itself is not significantly associated with the *Fried et al.* frailty phenotype in the mixed-sex population, some of the known frailty-related features such as WHR, ALM, morbidity index or polypharmacy manifest non-significant difference in either one of the three populations (Table 1). Even when significant for mixed-sex, men, or women, their effect sizes varied, suggesting that the strength of these associations with pre-frailty is larger in some groups than others. Similar observations have been made regarding the difference of these features across sex alone, or frailty alone (Fhon et al., 2018; Gutiérrez-Valencia et al., 2018; Karanth, 2023; Wang et al., 2024; Gordon et al., 2017; Uchai et al., 2023; Yarnall et al., 2017), but not yet in the combined context of sex-specific physical frailty. This could be because the resulting frailty phenotype, or the underlying frailty score, does not vary between men and women in the BASE-II (Fig. 2a-b). Instead, our results highlight the evidence of sex-specificity among the 5 *Fried et al.* frailty phenotype items (Fig. 2c-d, Table 2) that were used to define the level of frailty. As such, deficits in PA and GA are more prominent in pre-frail/frail men, while pre-frail/frail women more frequently report EX and show lower GS. In contrast, WL is equally represented in men and women, and double interactions of the sex-specific traits with WL do not follow this trend, which is likely due to the low number of self-reported cases of (unintended) WL. Besides the hallmarks of physical frailty, the *Fried et al.* frailty phenotype items, and their interactions, we also note clear variations between men and women in continuous biomedical data, independent of the frailty levels (Fig. 2e). These differences point

to sex-specific clinical manifestations: with men more affected by performance-related deficits (PA and GA), and women more affected by psychological or subjective components (EX), possibly reflecting the role of depression in women's frailty expression (Fhon et al., 2018; Gutiérrez-Valencia et al., 2018; Karanth, 2023; Wang et al., 2024; Gordon et al., 2017; Uchai et al., 2023; Yarnall et al., 2017).

3.3. Predictive models reveal divergent pre-frailty profiles in men and women

Our best-performing sex-specific models outperformed the mixed-sex model (Fig. 4a). sex-specific models more accurately predicted pre-frailty, while the mixed-sex model overpredicted non-frailty. Our analysis on the most clearly contributing clinical features of each model and the differences in technical configuration of these models confirms that men and women have different frailty trajectory profiles (Fig. 4c, Table 3). The mixed-sex model composed of these sex-specific profiles struggled to identify predictive features for men and women pre-frailty due to the asymmetric prevalence of most contributing features. Examples of such asymmetric features are vitamin D deficiency, sarcopenia, and alcohol consumption in women, and body fat mass, age, and height in men. While those features were identified in the mixed-sex model, they were only re-identified in one of the sex-specific contexts, aligning with several studies that also attributed distinct sex-specific importance to these features (Patel et al., 2017; Spira et al., 2015; Shardell et al., 2009; Spira et al., 2019). Among the top ten contributing features in all models, only one heart insufficiency showed symmetrical contributions regardless of sex. This feature is known to be highly associated with frailty, and independent of sex (Uchmanowicz et al., 2014; Sze et al., 2019). It is also noticeable that cognitive features, although not identical, ranked among the top ten features in all three models. While the digit symbol substitution test score was identified in the mixed-sex model to be lower in pre-frail/frail participants, both men and women models share the trail-making test as a predictive feature, which takes more time for pre-frail/frail participants to accomplish (coloured squares in Fig. 4c, for values see upper halves of Table 3, Tab. 3a-b). Although several studies have observed associations between poor CG and frailty (Karanth, 2023), the CG measurements included in the current analysis only poorly reflected the *Fried et al.* physical frailty (Fig. 4c, column 'CG- '). Cognitive frailty itself is a subtype of frailty and focuses explicitly on neurological pathways, markers, and CG decline in susceptible patients (Panza et al., 2015). It is important to emphasize that while these features are ranked by importance within each model, their individual predictive value in isolation is modest. Their utility lies in their combined contribution to model performance, and we avoid over-interpreting such marginal associations, as our findings suggest that the combined feature importance is greater than the sum of each individual feature importance.

3.4. Subgroup-based models suggest different biomedical pre-frailty signatures

In our second setup of prediction exercises, we focused on the various origins of biomedical data to identify the best performing subgroup in each context. Although this approach disables the comparability between contexts, it allows us to compare subgroup performance to the complete data-driven models within each context. The single subgroup performances (see heatmap, Fig. 5) tested for PM-, BF-, NT-, and CG-related features show minimal improvement in the mixed-sex population. However, performance for men significantly improved when limited to PM-related biomedical features, whereas the women model moderately improved when limited to BF-related features (Fig. 4b). This observation supports the idea that frailty manifests differently between sexes, suggesting the need for sex-specific subtyping of physical frailty rather than based on deficits such as dementia, sarcopenia, metabolic dysfunction, or a particular insufficiency (Rockwood and Mitnitski, 2007; Fried et al., 2021; Liu et al., 2017; Woo et al., 2015; Church et al., 2020). The top-contributing features measured for the best-performing subgroup data-driven mixed-sex and sex-specific models depicted in Fig. 4d indicate that especially the men model identified a strong frailty signal based on physical characteristics of the body, notably the trunk fat mass (p-value = 3.72e-04), followed by the ALM-BMI ratio (p-value = 3.65e-10). Moreover, the remaining features suggest a high fat and low lean mass profile of men frailty development (coloured squares in Fig. 4d, for values see lower half of Table 3a). This observation of distinct body composition in pre-frail/frail participants aligns with the current literature but has not been specifically shown to be more relevant in men than women (Uchai et al., 2023; Xu et al., 2020). The women model only slightly improved using BF-related features, with vitamin D deficiency contributing by far the most to the model's performance, with over 4 % of CV score loss (p-value = 5.96e-04). Additionally, white blood cell counts such as monocytes (p-value = 2.41e-02), eosinophiles (p-value = 4.30e-03), and leukocytes (p-value = 4.95e-03) appear to be higher in pre-frail/frail women (coloured squares in Fig. 4d, for values see lower half of Table 3b). Vitamin D deficiency and anomalies in blood composition both have recently gained more attention regarding their associations with frailty, although not in the sex-specific context as reported here (Shardell et al., 2009; Zhou et al., 2016; Spira et al., 2019; Mitchell et al., 2022). These results support the hypothesis of sex-specific physical frailty, with men more likely to exhibit an imbalanced physical shape, whereas pre-frail/frail women are more likely to present physiological anomalies or deficits, such as elevated white blood cell counts or vitamin D deficiency. These results suggest a potential sarcopenia-frailty link in men, as low ALM and fat mass-lean mass imbalance are top contributors. In contrast, pre-frailty in women appears more related to exhaustion and physiological factors such as inflammation and hormonal status, pointing to depression and immune dysregulation as possible drivers.

3.5. Feature interactions reinforce distinct sex-specific pre-frailty mechanisms

Finally, we investigated the interacting relationships of the top ten contributing features that best describe the dispersion between the non-frail and pre-frail/frail participants in *men-PM* and *women-BF* models using the MANOVA additive formulation of the single features in relation to the *Fried et al.* frailty phenotype (Fig. 5). Both the men (Fig. 5a) and women (Fig. 5b) model show increasing separation of non-frail and pre-frail/frail participants with every added feature up to all ten. The p-values for the various combinations of features peak for the *men-PM* model using the five following features: ALM-BMI ratio, ALM, whole body total BMD, left arm BMC, and left arm lean mass (MANOVA p-value = 2.36e-16). Noticeably, the single feature of ALM-BMI ratio shows a significant dispersion already (MANOVA p-value = 9.40e-15), and consequently this feature is included in all top combination of features (see table below Fig. 5a). Regarding the *women-BF* model, the best combination of features consists of vitamin D deficiency, oestradiol level, eosinophil count, HDL cholesterol level, and deoxypyridinoline in urine (MANOVA p-value = 9.21e-11). The best performing single feature remains vitamin D deficiency (MANOVA p-value = 1.40e-07), which is also included in all top combination of features (see table below Fig. 5b). With this approach, we substantiated the hypothesis that physical pre-frailty involves an accumulation of measurable deficits (Rockwood and Mitnitski, 2007; Lachmann et al., 2019), and that the nature of these deficits differs between sexes: in men, frailty aligns with traits of sarcopenia (ALM-BMI ratio, BMC, BMD), while in women, frailty aligns with hormonal insufficiency, immune activation, and exhaustion-related variables (white blood cell count, vitamin-D deficiency). Further analysis of the features' combinatory effect shows increasing separation of the pre-frail/frail from non-frail participants by the raw, PCA-, and LDA-transformed values with the one, two, five, and ten top features (Fig. 5), with the LDA version yielding overall the

highest BCD-WCD ratio. Although this simplified approach of combining and transforming features does not mimic the processing of our prediction pipeline (see Tab. S4 for details), it still reveals important insights into the sex-specific differences of the physical frailty profiles.

3.6. Sex-specific patterns in medication use and other frailty-associated features

In conclusion, we were able to identify sex-specific pre-frailty profiles in a cohort of individuals aged ≥ 60 years at the level of the *Fried et al.* 5-item frailty phenotype (Fig. 2), as well as at the level of currently known physical frailty-associated phenotypes (Fig. S1, Fig. S2, Table 1). Further analysis of this sex-specific divergence identified common, and more importantly, sex-specific frailty factors and patterns that subsequently improved the machine-learning-based prediction performance. Post-hoc analysis of the top-contributing features additionally revealed that pre-frail/frail participants are more likely to suffer from accumulated organ deficits and alcohol consumption behaviour as well as medication intake (Fig. S6, Fig. S7, Fig. S8). Concerning the latter, this suggests that pre-frail/frail women could require more medication due to possible higher disease burden or health-seeking behaviour. The opposite effect in men indicates that men who are less frail are taking more medications, potentially due to effective management, proactive healthcare, or a mere survivor effect. These contrasting patterns in medication use between men and women may also reflect sex-specific differences in comorbidity burden or treatment adherence, which have already been reported in literature showing such disparities in older adults' care and outcomes (Reid et al., 2022; Hessey et al., 2020; Dallmeier et al., 2020; Verschoor et al., 2024; Karanth, 2023; Wang et al., 2024). However, the individual features show only very minor differences between the healthy and pre-frail/frail participants (Fig. S10). This suggests that the machine-learning-based trained profile of pre-frail/frail participants is rather based on the interplay between the top contributing features. As such, our final exercise on the simple additive feature combinations was able to highlight the differences in men and women pre-frailty given their physical body shape and BF-based physiological profile, respectively (Fig. 5, Fig. S11).

3.7. Study limitations and conceptual framework considerations

There are several limitations to this study. During data cleaning, a substantial number of features with available data for at least 80 % of

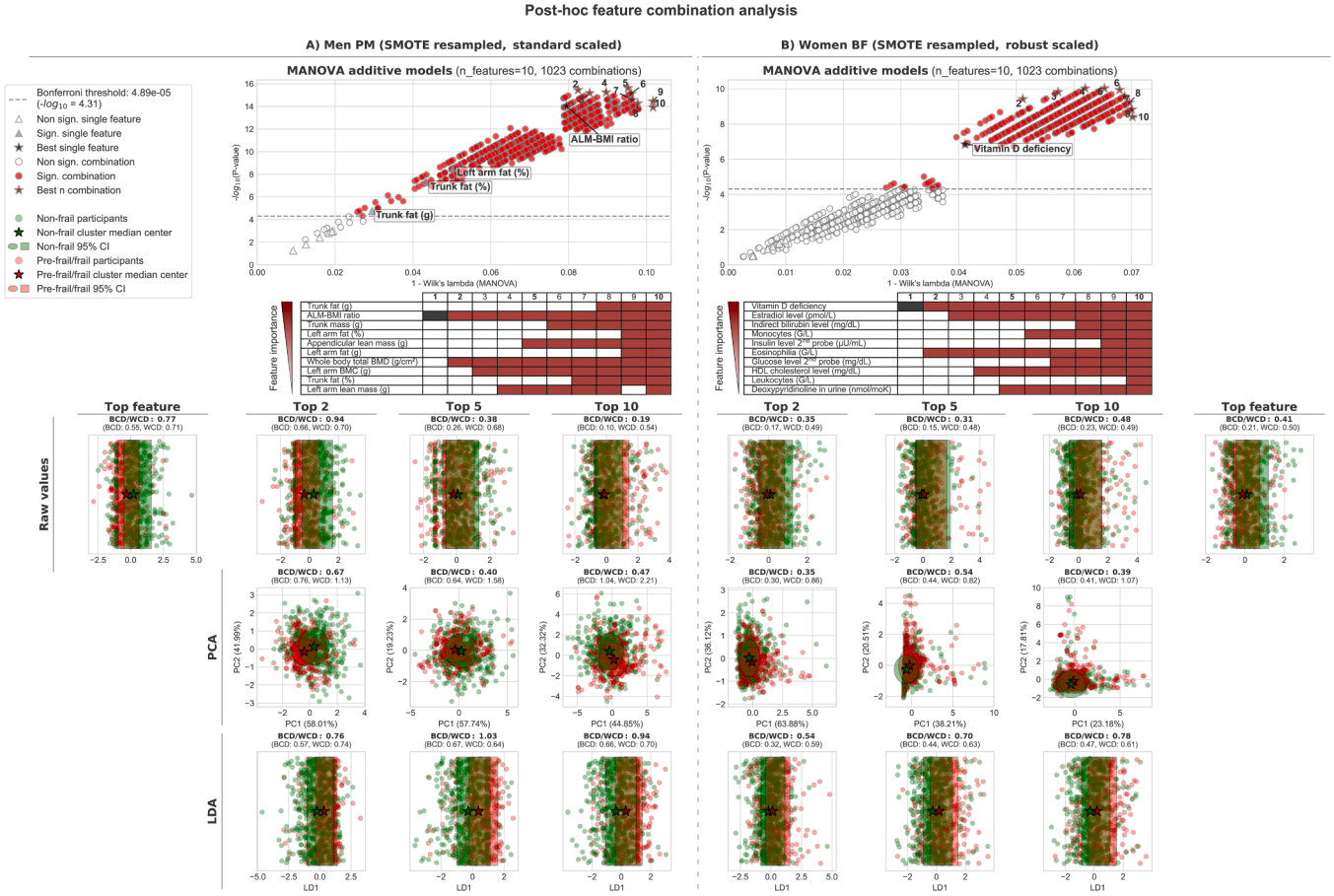


Fig. 5. Post hoc analysis of the best feature combinations shows increasing separation of the clusters between the best performing top single feature and the best combinations. The men (A) and women (B) subgroup data-driven models applied the respective sampling and scaling technique before undergoing MANOVA. Combinations from single to all ten features (1023 combinations in total) were tested and the performance was measured by the Wilk's lambda coefficient and represented as "1-Wilk's lambda". The resulting p-value was log base-10 transformed. Significant combinations are marked in red while the best performing set of n features are marked in dark red star shapes and annotated with the respective n. Single features are represented by triangles, with significant single features coloured in black. The table underneath the figure shows the top ten features in decreasing performance order and the components of the n best combinations are highlighted in dark red (black in case of single top significant feature). The p-value threshold of 0.05 was adjusted by the Bonferroni method and is visualized by the horizontal dashed line. Below the tables are the raw values, principal component analysis and linear discriminant analysis of the best single feature and the best combinations of two, five, and all ten features. The various representations are annotated with cluster distance characteristics to describe the separation and tightness of the non-frail (green) and frail (red) clusters. A 95 % CI is applied, and the cluster median centres are depicted in star shapes. BF body fluids, PM physical measurements, RUS random under-sampling, SMOTE synthetic minority over-sampling technique, sign. Significant, BCD between-cluster distance, WCD within-cluster distance.

participants needed imputation to retain as many participants as possible for the subsequent prediction analysis. We limited data leakage by imputing the train and test sets separately based on the statistical characteristics of the train set. Furthermore, the low number of participants affected by three or more *Fried et al.* frailty items (14 participants in total, 0.9 %) is a critical limitation to the statistical power and interpretation of findings related to full frailty. They have been merged with the pre-frail participants, potentially masking biomedical and pathophysiological constituents that could be able to describe not only the axis between non-frail and pre-frail/frail, but also the transition from pre-frail to frail in more detail. However, to mitigate the complete loss of these 'true' frail participants and their values, and to avoid resampling for this very small subset, pre-frail and frail participants were merged in one single category. Despite this, we were still able to investigate the performance of our binary-targeted models on the three original *Fried et al.* labels and showed that in fact the models were able to predict nearly all frail and most pre-frail participants correctly (Fig. 6). Here, the identified predictors in the combined pre-frail/frail models remain significantly predictive in the much smaller frail subgroup with increased sensitivities compared to the pre-frail subgroup. However, as the merged category is dominated by pre-frail participants, the present results reflect biomarkers of pre-frailty, not frailty. Pre-frail and frail are distinct clinical stages, and future studies with larger frail cohorts will be required to extend these findings to fully developed frailty. Another machine-learning-based limitation is the cross-validated scoring system in the case of imbalanced datasets, as after merging pre-frail and frail, there are still twice as many non-frail participants in the cleaned dataset. We addressed these limitations by selecting a validation score that prioritizes recall over precision, namely the F-β-2 score (see Eq. 2) and considered under- and over-sampling techniques during the training process to mitigate the data imbalance in BASE-II. Still, since the available biomedical feature space spans predictors that can be linked to *Rockwood*'s frailty index and accumulation of deficit model (Rockwood et al., 1994; Rockwood and Mitnitski, 2007), we acknowledge this conceptual discrepancy as potential source of overfitting or model bias, particularly when interpreting biomedical contributors to the *Fried et al.* frailty phenotype. We therefore framed our findings as exploratory and hypothesis-generating. Another major limitation of this study is the cross-sectional nature of the biomedical data used in the current analysis, meaning that they only reflect a snapshot of the participants. Having access to multiple cross-sections at different timepoints would be important and interesting, e.g., it could help to investigate whether the false-positive participants would later develop frailty or not. Moreover importantly, our models were only validated using cross-validation techniques within the BASE-II cohort, and external validation in an independent dataset is needed to confirm generalizability of the presented predictors. Finally, the analysis of the top feature combinations as shown with MANOVA is not exactly representing the model behaviour as it did during the training process. Instead, it is a simpler, linear, and independent representation of the combined features' power to separate healthy from pre-frail/frail participants by simple addition of their scaled values, PCA-driven, and LDA-driven data transformation (Fig. 5). Altogether, further research is required to better understand the sex-specific differences in pre-frailty that we have thoroughly observed and analysed in BASE-II.

4. Materials and methods

4.1. BASE-II participants and frailty phenotype

BASE-II is an observational study including 2200 participants from age 20–35 and 60–85, recruited between 2009 and 2015 in the greater Berlin area, Germany (see https://www.base2.mpg.de/en, and references (Bertram et al., 2014; Baltes and Mayer, 1998; Demuth et al., 2021). All BASE-II participants provided written informed consent before participation, and the study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Charité – Universitätsmedizin Berlin (approval number EA2/029/09; date of approval: 19 March 2009). Participants attended a wide range of examinations resulting in the rich collection of psychological, genetic, medical, socioeconomic, and immunological data. Frailty in BASE-II was measured by the 5-item *Fried et al.* frailty phenotype (Fried et al., 2001), composed of the clinical variables WL, PA, EX, GS, and GA, with adjustments described in Spira et al (Spira et al., 2015). (see Tab. S1). The presence of frailty was determined with the below Eq. (1) as the sum of positive frailty-related phenotypes to give a frailty score (FR_{score}).

$$FR_{score} = WL + PA + EX + GS + GA \tag{1}$$

The *Fried et al.* frailty phenotype is derived from the resulting FR_{score} and participants are labelled as either non-frail (level 0, $FR_{score} = 0$), pre-frail (level 1, $FR_{score} \in \{1;2\}$), or frail (level 2, $FR_{score} \geq 3$).

4.2. Data cleaning

In this study, a cross-section of the BASE-II cohort was analysed with a focus on the clinical and socioeconomic characteristics of the older study population. Data cleaning included the removal of features with low data coverage and participants missing the information to determine the frailty phenotype, engineering of features for clinically relevant characteristics linked to frailty risk factors (i.e., vitamin D

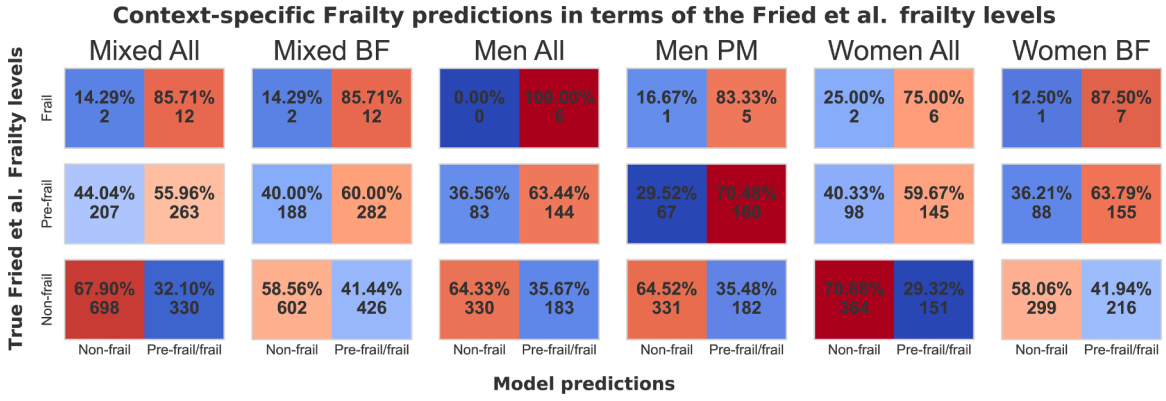


Fig. 6. Context-specific frailty predictions in terms of the *Fried et al.* frailty phenotype levels. Each model's prediction performance of true frail, pre-frail, and non-frail are shown as number and percentage of the total context-specific BASE-II population, combining the training and test set participants. The best *mixed-BF* model yields sensitivities of 60.0 % in the pre-frail and 85.7 % in the frail subgroup, with a false positive rate (FPR) of 41.4 % and specificity of 58.9 %. The best *men-PM* model yields sensitivities of 70.5 % in the pre-frail and 83.3 % in the frail subgroup, with a FPR of 35.5 %, and specificity of 64.5 %. The best *women-BF* model yields sensitivities of 63.8 % in the pre-frail and 87.5 % in the frail subgroup, with a FPR of 41.9 %, and specificity of 58.1 %.

deficiency (Shardell et al., 2009; Zhou et al., 2016; Spira et al., 2019), polypharmacy including reported drug side-effects, allergies, and drug-specific disease area (Gutiérrez-Valencia et al., 2018; Nwadiugwu, 2020; König et al., 2018), and poly-dependency (Abreu et al., 2019)); the labelling of feature subgroups (10 subgroups: individual medications (IM), individual devices (ID), grouped medications (GM), surveys (SV), chronic morbidity (CH), grouped devices (GD), NT, PM, BF, and CG (Fig. 3a); splitting into training and test set; and imputation of the remaining missing values. With the scope of this study being the prediction of pre-frailty in older people, and the literature especially showing phenotypic relationships between older age groups and frailty (Lally and Crome, 2007; Clegg et al., 2013; Fulop et al., 2010; Rockwood et al., 1994; Abizanda et al., 2016; Holliday, 1995; Fedarko, 2011; Fhon et al., 2018), participants of age 60 or above were retained. Of the initial 2171 participants with baseline data assessed in the medical part of BASE-II, 1671 satisfy this criterion.

4.2.1. Handling missing values and redundancies

In general, features with more than 20 % of missing values are considered not informative enough to describe the entire dataset for the task of training a prediction model (Broeck et al., 2005; Eekhout et al., 2012). Of the initial 975 raw BASE-II features, 555 features had more than 20 % of missing values. Among those features, 506 binary variables are related to specific medication intake or using a specific helping device (1 = yes, 0 = no). Due to the large number of such features and their sparsity (most participants reported using only a few), we later aggregated them into counts of total helping devices and total medication per target disease (see Section 4.2.3). For these features, missing values were likely the result of non-response rather than actual usage, and we therefore treated missing values as negative response. In contrast, 40 other features related to clinical walking assessments were missing for most participants. The disruption in these features due to changes in the methodology, leaving them unusable for classification. The remaining 9 features with low coverage were redundant and thus discarded. With 919 remaining features having missing value coverage of less than 20 %, one additional feature related to sex was removed due to redundancy, being available both in Boolean (0 =men, 1 =women) and discrete form (1 =men, 2 =women). Furthermore, 159 participants were missing one or multiple frailty-related phenotypes necessary for Eq. (1) and were therefore removed. After handling the features with missing values and redundancies in BASE-II, the size of the dataset was reduced to 1512 participants and 918 features.

4.2.2. Feature engineering

Feature engineering was performed to reduce the sparsity within the IM and ID features, and to combine the available measurements for two different methodologies of measuring 25-hydroxy-vitamin D levels in serum conducted in BASE-II. The IM subgroup constituted 452 features in total, reflecting various information on 124 drugs, including name, administered dose, active substance, and experienced side-effects. Like the approach of König et al. (König et al., 2018), we engineered the raw medication intake counts as a new feature called polypharmacy (ranging from 0 to 15) and summed up the number of experienced side-effects and allergies for each participant as cumulative side-effects (ranging from 0 to 3) and cumulative allergies (ranging from 0 to 2).

Furthermore, the drug-related disease area of the 124 drugs described in BASE-II was determined using the Drug Repurposing Hub of the Broad Institute (version of 24th May 2021, see https://clue.io/repurposing-app, and reference (Corsello et al., 2017)), a curated collection of more than 6700 drugs, either FDA-approved or in clinical trial, as well as the Drug Information Database Drugs.com (see https://www.drugs.com/drug_information.html, and reference (Leah Plumb, 2004)). With all correct FDA-approved names at hand, we extracted 93 unique drug interactions, 77 unique modes of action, and 18 disease areas. To effectively reduce the number of features related to IM, only the targeted disease area was retained (cardiology, dental, dermatology, endocrinology, gastroenterology, haematology, infectious disease, metabolism, nephrology, neurology/psychiatry, obstetrics/gynaecology, oncology, ophthalmology, orthopaedics, otolaryngology, pulmonary, rheumatology, and urology). The number of drugs targeting the same disease area was computed for each participant, ranging up to a maximum of 7 drugs (cardiology). Besides IM-derived polypharmacy, an identical approach was used to transform the ID information of the BASE-II participants (including 55 distinct assistive or helping devices) into the single feature poly-dependency, shown previously to be linked with medication intake, overcoming physical disabilities, moderate or severe frailty, and dementia (Abreu et al., 2019; Tomita et al., 2004).

Then, two low-coverage features related to 25-hydroxy-vitamin D levels in blood were identified in BASE-II with complementary missing values, representing two distinct measuring methods conducted during the study. As both features showed similar distributions, they were merged before computing the standard z-score, creating a new normally distributed feature of vitamin D levels. In addition to the merged z-score values, another binary feature representing vitamin D deficiency was engineered based on 25-hydroxy-vitamin D levels according to the recommendations of the Institute of Medicine (Spira et al., 2019; Ross et al., 2011), reporting 50 nmol/L or less as a threshold for vitamin D deficiency.

Lastly, the feature of interest, frailty phenotype, was scaled down from the three initial levels non-frail ($FR_{score} = 0$; $N = 1028$), pre-frail ($2 \geq FR_{score} \geq 1$; $N = 470$), and frail ($FR_{score} \geq 3$; $N = 14$) to only two levels, the non-frail and pre-frail/frail group. Latter combines both pre-frail and frail, totalling 484 participants. This combination was necessary due to a pronounced imbalance between the three groups, a circumstance that could potentially exert a detrimental influence on the performance of machine-learning-based classification. Importantly, since this merged category consists almost entirely of pre-frail individuals, results of this study should be interpreted as predictors of pre-frailty. No reliable inference about fully developed frailty is possible from this data.

Following feature engineering, the BASE-II dataset dimensions increased from 918 to 942 by 21 GM (side-effects, allergies, multiple disease areas and polypharmacy), one GD (poly-dependency), and two vitamin D-related engineered features (vitamin D level z-score and vitamin D deficiency status).

4.2.3. Stratified data splitting and imputation

Stratified splitting of the reshaped dataset was performed generating a training and hold-out test set with an 80-to-20 ratio using MATLAB's Statistics and Machine-learning Toolbox *cvpartition* function with the hold-out method. This method ensures to preserve the same class distribution in both sets and prevents bias in the model due to imbalanced class proportions (non-frail to pre-frail/frail ratio in complete BASE-II: 2.12, training: 2.18, test: 1.90). Of the 1512 participants (746 men, 766 women), 1210 composed the training set (593 men, 617 women), and 302 the test set (153 men, 149 women). Imputation was then performed in both training and test sets based on the information from the training set only. Discrete and binary features in the training and test set were imputed using the mode, while continuous features were imputed using the mean of the training set given their non-skewed distributions (Amballa, 2020).

4.3. Data processing

After the data has been cleaned and imputed, cohort characteristics were first computed for the most important frailty-related variables according to the literature, including sex (Spira et al., 2019; Gordon et al., 2017), age (Fulop et al., 2010; Fedarko, 2011), ALM, ALM-to-BMI ratio and sarcopenia (Spira et al., 2015; Fhon et al., 2018; Uchai et al., 2023); as well as heart insufficiency (Uchmanowicz et al., 2014; Sze et al., 2019), vitamin D deficiency (Shardell et al., 2009; Zhou et al., 2016; Spira et al., 2019), morbidity index (Espinoza et al., 2018; Tenison

and Henderson, 2020; Yarnall et al., 2017), and polypharmacy (Gutiérrez-Valencia et al., 2018; Nwadiugwu, 2020; König et al., 2018). Subsequently, the dataset was cleared from features used for feature engineering then split by sex and partitioned by the subgroups that have the most available information (PM, NT, CG, BF). Furthermore, we removed constant and highly correlated features within each sex- and context-specific combination. This data processing part is the first main step of the pipeline we have created for this study (see https://github.com/sysbiolux/Clinical_Biomarker_Detection) and is an indispensable step before applying machine-learning-based strategies for classification in the second step. Data processing and machine-learning were performed using the python programming language and the High-Performance Computing (HPC) facilities of the University of Luxembourg (see https://hpc.uni.lu/, and reference (Varrette et al., 2022)).

4.3.1. BASE-II cohort characteristics

In Table 1, we conducted comparative analysis of the cohort characteristics between the non-frail and pre-frail/frail participants in mixed-sex, men, and women. The statistical significance between each group was computed using the Welch's unequal variance T-test (WELCH, 1947) (with the equality of variance determined by the Levene's test) for continuous features, and the Cramér's V correlation algorithm (Cramér, 1946) corrected for binary and discrete feature data types (Bergsma, 2013). The characteristics are either presented in mean and standard deviation (continuous features), or the counts of the categories (binary and discrete), and their interquartile range (discrete).

4.3.2. Data filtering and partitioning

Of the 942 cleaned and imputed cohort features, the two sparse subgroups IM ($N = 453$) and ID ($N = 55$) were taken care of during the feature engineering and thus were removed from the dataset. Additionally, seven features related to the *Fried et al.* frailty phenotype were removed as they directly or indirectly represent the five frailty characteristics WL, PA, GS, GA, and EX, as well as the resulting FR_{score}. The reduced dataset is further separated by sex, resulting in three distinct datasets: mixed-sex, men, and women. While each dataset underwent a classification exercise using all remaining subgroups, denoted 'ALL', we additionally conducted context-specific experiments for the largest subgroup datasets, namely BF, PM, NT, and CG ($N \geq 40$), resulting in 15 distinct models (Fig. 3a-b).

4.3.3. Removing constant and near-constant features

A first round of quality checks of the partitioned datasets consisted of the identification of strictly constant (identical) features or near-constant (extremely low variance-to-mean ratio or positive cases) features (see Fig. 3c). For each analytical context (mixed-sex, men, women), only a very few strictly constant features were identified and thus removed. Considering near-constant features, two distinctions were made to identify them either within the continuous features or the binary and categorical features. In the realm of continuous features, two arbitrary thresholds were considered to define near-constancy by using the variance-to-mean ratio of 0.01 and 0.001 as cut-off. With this, only features were kept that, considering their mean, show at least a 1 % or 0.1 % variance to it within the population. All datasets and partitions thereof were assessed for both suggested thresholds, and the best performing thresholds were retained. Near-constant binary or categorical features were removed if the lowest class count is below the number of CV splits applied in the machine-learning part, which was set to 10 splits. A more detailed summary of the number of constant and near-constant features that have been removed in each model can be reviewed in the supplementary files (Tab. S2).

4.3.4. Removing highly correlated features

A subsequent round of quality check of the datasets and partitions thereof aimed to identify and remove features that are highly correlated to each other based on predefined correlation coefficient thresholds for every possible data-type association. Not only was the dimensionality of each dataset further reduced by this step, but it was also cleaned of redundant features that would not necessarily bring additional information to the models but rather make it more complex and more prone for the risk of errors. Three different data-type associations were considered, and their correlations computed appropriately: the continuous-continuous associations using the Spearman rank coefficient (Spearman, 1904), the categorical-categorical associations using the Cramér's V correlation coefficient corrected for binary and discrete feature, and the continuous-categorical associations using the point bi-serial correlation coefficient (Glass and Hopkins, 1996). For each of the three associations, multiple thresholds were tested and the thresholds leading to the best performing downstream classifications were retained (see Fig. 3d-e, and supplementary figures Fig. S3a-b). For continuous-categorical associations, the features removed were always selected from the group of continuous features. The number of features removed in each category can be reviewed in the supplementary files (Tab. S2).

4.4. Machine-learning

The second main part of the pipeline is dedicated to the machine-learning-based classification. Due to the imbalance of frailty in BASE-II, the datasets were resampled based on two distinct approaches: RUS and SMOTE (Chawla et al., 2002). Then, feature scaling and dimensionality reduction were applied. We assessed three scaling strategies (min-max, standard, and robust scaler) and retained for each dataset the strategy that contributed to the best prediction performance of frailty. Reducing the dimensionality was performed using the two linear approaches of PCA and LDA, as well as the non-linear kernel PCA for continuous features, whereas discrete features were directly selected using the k best features correlated to the frailty phenotype target. Therefore, we assessed two different correlation statistics, namely chi-squared (χ^2) and Cramér's V correlation coefficient corrected for binary and discrete data types. As the order of manipulations (resampling and feature transformation) can affect the prediction performance and no preferred order is given for the scenario of imbalance classification (Zhang et al., 2017), we assessed both possibilities and observed that first resampling followed by feature transformation was always the more performant strategy to predict frailty.

The resampled and transformed datasets were then passed to a SVM classifier (Boser et al., 1992) using linear and non-linear kernels. The best-fitting model parameters in each scenario were identified using the exhaustive CV grid search method for large ranges of hyper-parameter values (Fig. 3**e**). Model selection was based on the best CV F-β-2 score during training, which emphases recall twice as much as precision (see Eq. 2), making it particularly suitable for tasks where minimizing false negatives is a high priority in imbalanced datasets.

$$F_\beta = \left(1 + \quad \beta^2\right) * \frac{precision * recall}{\left(\beta^2 \quad * precision\right) + recall} \tag{2}$$

Each model was evaluated using multiple metrics during the CV, the training process, and within the hold-out validation test set (Fig. 3f). These metrics include the ROC, diagnostics odds ratio (DOR), true positives (TP), and true negatives (TN). From the best models, feature contributions were assessed using feature permutation, which measures the difference in model CV training score when single feature values are randomly permuted to mimic a perturbation in the machine-learning based prediction system (Breiman, 2001). A detailed summary of the minimum information for medical artificial intelligence reporting (see (Hernandez-Boussard et al., 2020)) can be found in Tab. S3.

4.4.1. Hyper-parameter optimization and stratified k-fold cross-validation

Hyper-parameter optimization and stratified k-fold CV were

performed using exhaustive grid search of preselected parameter ranges (see Tab. S4) using stratified 10-fold CV while training, and the F-β-2 scorer (see Eq. 2). For each partition of the data (by sex and by feature subgroups), we documented the most suitable models and restricted their optimal parameters to a more limited range. This process of constraint iteration was repeated up to four times, depending on whether it could lead to enhanced performance of the model. For the combined subgroups model, the best performance was achieved after one single iteration in the case of the mixed-sex and men model, after the third one in the women model, and the fourth one in the other subgroup-specific models.

4.4.2. Measuring feature importance

For each optimized model, feature importance was assessed using systematic feature permutation, in which features were randomized individually at least 500 times, and the resulting CV score are compared to the original ones. The contribution of each feature is represented by the mean and standard deviation of change of the optimized CV score after shuffling. Feature robustness was determined via z-score. We limited the analysis of the best contributing features to only the top thirty features with the largest change in CV score. Importantly, permutation scores reflect the relative significance of individual feature in the model, not their standalone predictive power.

4.4.3. Subsequent feature extractions that improved sex-specific models

Next, we further investigated the top features for each model. The top thirty features were reused as input features, and two to three context-specific hallmarks were further added if they were not already within these thirty features. This approach improved the prediction performance in specific contexts: subgroup-driven men model using PM, and all-driven women models using all clinical subgroups combined. In the PM subgroup-driven men model, the raw ALM in grams and kilograms were added to the thirty most contributing features and showed a higher performance using the same information in two different units. For the all-driven women model, raw ALM in kilograms, vitamin D deficiency and vitamin D levels z-score were added. These models also underwent hyper-parameter optimization to constraint the parameters and further enhance the performance, which was observed after two iterations in both cases.

4.4.4. Multi-variate model evaluation

All hyper-parameterisation exercises used the F-β-2 scorer to evaluate the stratified 10-fold CV and select the best fitting model. Although this scorer prioritizes recall over precision, it does not guarantee that the model with the highest F-β-2 score has a better recall than precision. As this was observed in early experiments. To address this, we supplemented the CV evaluation with independent five times 10-fold CV ROC and precision-recall (PR) curves to better assess the training performance. In addition to CV, F-β-2, ROC, and PR were measured for the overall training performance and the hold-out validation test set. DOR, F-β-1, and accuracy were also measured in the overall training and test set. TP and TN were recorded from the hold-out test set as well.

4.4.5. Post-hoc feature contribution analysis

The ten most contributing features for each model were analysed for their individual and combined influence on the distinction between non-frail and pre-frail/frail participants in the training set. This included deficits (vitamin D deficiency, heart insufficiency, and sarcopenia) in the mixed-sex model, medication intake (ophthalmology and gastroenterology) in the men model, and alcohol intake in the women model. In each case, the proportion of pre-frail/frail participants is represented in the respective figures. Associations between the individual top continuous features and the frailty phenotype were analysed in a combined error bar and letter-value plot for the non-frail and pre-frail/frail mixed-sex, men, and women. The mean z-score difference between pre-frail/frail and non-frail participants of all available continuous features was plotted and significant top ten features were highlighted. For the non-continuous features, the Cramér's V correlation coefficient corrected for binary and discrete data types was plotted with a significance threshold of 0.10. In both exercises the p-value thresholds were adjusted using the Bonferroni method. Feature combinations were then analysed using the linear MANOVA method (formula: "$feature_1 + feature_2 + \ldots + feature_n \sim target$"), reporting Wilk's lambda coefficient as well as the best combinations ($1 \leq n \leq 10$). Given the multifactorial nature of frailty, this post-hoc analysis allowed us to evaluate the collective contribution of selected features, aiming to identify combinations with enhanced discriminatory power beyond what univariate analyses could capture (Beam et al., 2015). In the examples of the *men-PM* and *women-BF* models, the resulting combinations of the best two, five, and all ten features were plotted as raw values and in the form of reduced dimensions using PCA and LDA to showcase how well they separate non-frail from pre-frail/frail participants.

CRediT authorship contribution statement

Schneider Jochen G: Writing – review & editing, Formal analysis, Conceptualization. **Ali Kishk:** Writing – review & editing, Conceptualization. **Maria Pires Pacheco:** Writing – review & editing, Conceptualization. **Sébastien De Landtsheer:** Writing – review & editing, Formal analysis, Conceptualization. **Ilja Demuth:** Writing – review & editing, Data curation, Conceptualization. **Dominik Spira:** Writing – review & editing, Data curation, Conceptualization. **Graham Pawelec:** Writing – review & editing, Data curation, Conceptualization. **David Goldeck:** Writing – review & editing, Data curation, Conceptualization. **Jeff Didier:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Thomas Sauter:** Writing – review & editing, Supervision, Funding acquisition, Formal analysis, Conceptualization.

Author contributions

All authors have contributed to the study conception and design. Material preparation was done by JD, data contribution by DG, GP, DS, and ID, and analysis by JD, SDL, JGS, and TS. The first draft of the manuscript was written by JD with contributions from SDL, AK, and MPP. All authors read and approved the final manuscript for submission.

Declaration of Competing Interest

No financial or non-financial conflicts of interest have been identified for disclosure purposes.

Acknowledgments

The experiments presented in this paper were carried out using the High-Performance Computing (HPC) facilities of the University of Luxembourg (see https://hpc.uni.lu/, and reference (Varrette et al., 2022)). The authors would like to thank the steering committee of BASE-II and the BASE-II participants.

The Doctoral Training Unit Data-driven computational modelling and applications (DRIVEN) is funded by the Luxembourg National Research Fund under the PRIDE program (PRIDE17/12252781). This article uses data from the Berlin Aging Study II (BASE-II) which was supported by the German Federal Ministry of Education and Research under grant numbers #01UW0808; #16SV5536K, #16SV5537, #16SV5538, #16SV5837, #01GL1716A and #01GL1716B.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.mad.2025.112114.

Data availability

The authors do not have permission to share data.

References

Abizanda, P., Romero, L., Sánchez-Jurado, P.M., Ruano, T.F., Ríos, S.S., Sánchez, M.F., 2016. Energetics of aging and frailty: the FRADEA study. J. Gerontol. Ser. A. 71 (6), 787–796.

Abreu, W., Tolson, D., Jackson, G.A., Staines, H., Costa, N., 2019. The relationship between frailty, functional dependence, and healthcare needs among community-dwelling people with moderate to severe dementia. Health Soc. Care Commun. 27 (3), 642–653.

Akbari, G., Nikkhoo, M., Wang, L., Chen, C.P.C., Han, D.S., Lin, Y.H., et al., 2021. Frailty level classification of the community elderly using microsoft Kinect-Based skeleton pose: a machine learning approach. Sensors 21 (12), 4017.

Amballa, A., 2020. Feature engineering Part-1 mean/ median imputation. Analytics Vidhya. ⟨https://medium.com/analytics-vidhya/feature-engineering-part-1-mean-median-imputation-761043b95379⟩.

Baltes, P.B., Mayer, K.U. (Eds.), 1998. The Berlin Aging Study: Aging from 70 to 100 [Internet]. Cambridge University Press, Cambridge. ⟨https://www.cambridge.org/core/books/berlin-aging-study/B1641E71F86F59A1A20BBE8824CB2EE5⟩.

Beam, A.L., Motsinger-Reif, A.A., Doyle, J., 2015. An investigation of gene-gene interactions in dose-response studies with Bayesian nonparametrics. BioData Min. 6.

Bergsma, W., 2013. A bias-correction for Cramér's V and Tschuprow's T. J. Korean Stat. Soc. 42 (3), 323–328.

Bertram, L., Böckenhoff, A., Demuth, I., Düzel, S., Eckardt, R., Li, S.C., et al., 2014. Cohort profile: the Berlin aging study II (BASE-II). Int. J. Epidemiol. 43 (3), 703–712.

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory [Internet]. Association for Computing Machinery, New York, NY, USA, pp. 144–152 (COLT '92). Available from: https://dl.acm.org/doi/10.1145/130385.130401.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Broeck, J.V., den, Cunningham, S.A., Eeckels, R., Herbst, K., 2005. Data cleaning: detecting, diagnosing, and editing data abnormalities. PLOS Med. 2 (10), e267.

Buchmann, N., Spira, D., König, M., Demuth, I., Steinhagen-Thiessen, E., 2019. Frailty and the metabolic syndrome — results of the Berlin aging study II (BASE-II). J Frailty Aging 8 (4), 169–175.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Int. Res. 16 (1), 321–357.

Chen, X., Mao, G., Leng, S.X., 2014. Frailty syndrome: an overview. Clin. Interv. Aging 9, 433–441.

Church, S., Rogers, E., Rockwood, K., Theou, O., 2020. A scoping review of the clinical frailty scale. BMC Geriatr. 20 (1), 393.

Clegg, A., Young, J., Iliffe, S., Rikkert, M.O., Rockwood, K., 2013. Frailty in elderly people. Lancet 381 (9868), 752–762.

Collard, R.M., Boter, H., Schoevers, R.A., Oude Voshaar, R.C., 2012. Prevalence of frailty in community-dwelling older persons: a systematic review. J. Am. Geriatr. Soc. 60 (8), 1487–1492.

Corsello, S.M., Bittker, J.A., Liu, Z., Gould, J., McCarren, P., Hirschman, J.E., et al., 2017. The drug repurposing hub: a next-generation drug library and information resource. Nat. Med. 23 (4), 405–408.

Cramér, H., 1946. Mathematical methods of statistics. Princeton University Press, p. 598.

Dallmeier, D., Braisch, U., Rapp, K., Klenk, J., Rothenbacher, D., Denkinger, M., et al., 2020. Frailty index and sex-specific 6-year mortality in community-dwelling older people: the ActiFE study. J. Gerontol. A Biol. Sci. Med Sci. 75 (2), 366–373.

Demuth, I., Banszerus, V., Drewelies, J., Düzel, S., Seeland, U., Spira, D., et al., 2021. Cohort profile: follow-up of a Berlin aging study II (BASE-II) subsample as part of the GendAge study. BMJ Open 11 (6), e045576.

Di Sabatino, A., Lenti, M.V., Cammalleri, L., Corazza, G.R., Pilotto, A., 2018. Frailty and the gut. Dig. Liver Dis. 50 (6), 533–541.

Dibello, V., Zupo, R., Sardone, R., Lozupone, M., Castellana, F., Dibello, A., et al., 2021. Oral frailty and its determinants in older age: a systematic review. Lancet Healthy Longev. 2 (8), e507–e520.

Eekhout, I., de Boer, R.M., Twisk, J.W.R., de Vet, H.C.W., Heymans, M.W., 2012. Missing data: a systematic review of how they are reported and handled. Epidemiology 23 (5), 729.

Espinoza, S.E., Quiben, M., Hazuda, H.P., 2018. Distinguishing comorbidity, disability, and frailty. Curr. Geriatr. Rep. 7 (4), 201–209.

Fedarko, N.S., 2011. The biology of aging and frailty. Clin. Geriatr. Med. 27 (1), 27–37.

Fhon, J.R.S., Rodrigues, R.A.P., Santos, J.L.F., Diniz, M.A., dos Santos, E.B., Almeida, V. C., et al., 2018. Factors associated with frailty in older adults: a longitudinal study. Rev. Saúde Pública 52, 74.

Fried, L.P., Tangen, C.M., Walston, J., Newman, A.B., Hirsch, C., Gottdiener, J., et al., 2001. Frailty in older adults: evidence for a phenotype. J. Gerontol. Ser. A 56 (3), M146–M157.

Fried, L.P., Cohen, A.A., Xue, Q.L., Walston, J., Bandeen-Roche, K., Varadhan, R., 2021. The physical frailty syndrome as a transition from homeostatic symphony to cacophony. Nat. Aging 1 (1), 36–46.

Fulop, T., Larbi, A., Witkowski, J.M., McElhaney, J., Loeb, M., Mitnitski, A., et al., 2010. Aging, frailty and age-related diseases. Biogerontology 11 (5), 547–563.

Glass, G.V., Hopkins, K.D., 1996. Statistical methods in education and psychology [Internet]. Allyn and Bacon, Boston, p. 698. ⟨http://archive.org/details/statisticalmetho00glas⟩ (Available from).

Gomez-Cabrero, D., Walter, S., Abugessaisa, I., Miñambres-Herraiz, R., Palomares, L.B., Butcher, L., et al., 2021. A robust machine learning framework to identify signatures for frailty: a nested case-control study in four aging european cohorts. GeroScience 43 (3), 1317–1329.

Gordon, E.H., Peel, N.M., Samanta, M., Theou, O., Howlett, S.E., Hubbard, R.E., 2017. Sex differences in frailty: a systematic review and meta-analysis. Exp. Gerontol. 89, 30–40.

Gutiérrez-Valencia, M., Izquierdo, M., Cesari, M., Casas-Herrero, Á., Inzitari, M., Martínez-Velilla, N., 2018. The relationship between frailty and polypharmacy in older people: a systematic review. Br. J. Clin. Pharmacol. 84 (7), 1432–1444.

Hernandez-Boussard, T., Bozkurt, S., Ioannidis, J.P.A., Shah, N.H., 2020. MINIMAR (MINimum Information for Medical AI Reporting): developing reporting standards for artificial intelligence in health care. J. Am. Med Inf. Assoc. 27 (12), 2011–2015.

Hessey, E., Montgomery, C., Zuege, D.J., Rolfson, D., Stelfox, H.T., Fiest, K.M., et al., 2020. Sex-specific prevalence and outcomes of frailty in critically ill patients. J. Intensive Care 8 (1), 75.

Holliday, R., 1995. Understanding ageing [Internet]. Developmental and Cell Biology Series. Cambridge University Press, Cambridge. ⟨https://www.cambridge.org/core/books/understanding-ageing/7C795DA1F52124A30E0F6A6278A4C8E8⟩.

Howlett, S.E., Rutenberg, A.D., Rockwood, K., 2021. The degree of frailty as a translational measure of health in aging. Nat. Aging 1 (8), 651–665.

Idris S., Badruddin N. Classification of Cognitive Frailty in Elderly People from Blood Samples using Machine Learning. In: 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI). 2021. p. 1–4.

Karanth, S.D., 2023. Sex-specific association of physical frailty and cognitive function in a population-based cross-sectional study of American older adults. Alzheimers Dement 19 (S6), e068336.

Koenig M., Malsch C., Marino J., Vetter V.M., Komleva Y., Demuth I., et al. Nocturia as a Risk Factor for Developing Frailty in Older Adults: Results of the Berlin Aging Study II [Internet]. medRxiv; 2024 [cited 2024 Sep 24]. p. 2024.09.20.24313292. Available from: ⟨https://www.medrxiv.org/content/10.1101/2024.09.20.24313292v1⟩.

König, M., Spira, D., Demuth, I., Steinhagen-Thiessen, E., Norman, K., 2018. Polypharmacy as a risk factor for clinically relevant sarcopenia: results from the Berlin aging study II. J. Gerontol. Ser. A 73 (1), 117–122.

Lachmann, R., Stelmach-Mardas, M., Bergmann, M.M., Bernigau, W., Weber, D., Pischon, T., et al., 2019. The accumulation of deficits approach to describe frailty. PLOS ONE 14 (10), e0223449.

Lally, F., Crome, P., 2007. Understanding frailty. Post. Med. J. 83 (975), 16–20.

Laur, C.V., McNicholl, T., Valaitis, R., Keller, H.H., 2017. Malnutrition or frailty? Overlap and evidence gaps in the diagnosis and treatment of frailty and malnutrition. Appl. Physiol. Nutr. Metab. 42 (5), 449–458.

Leah Plumb, A., 2004. Drugs.com: drug information online 2004. Ref. Rev. 18 (6), 41–41.

Leme, D.E., da, C., de Oliveira, C., 2023. Machine learning models to predict future frailty in community-dwelling middle-aged and older adults: the ELSA cohort study. J. Gerontol. A Biol. Sci. Med. Sci. 78 (11), 2176–2184.

Linn, N., Goetzinger, C., Regnaux, J.P., Schmitz, S., Dessenne, C., Fagherazzi, G., et al., 2021. Digital health interventions among people living with frailty: a scoping review. J. Am. Med. Dir. Assoc. 22 (9), 1802–1812 e21.

Liu, L.K., Guo, C.Y., Lee, W.J., Chen, L.Y., Hwang, A.C., Lin, M.H., et al., 2017. Subtypes of physical frailty: latent class analysis and associations with clinical characteristics and outcomes. Sci. Rep. 7 (1), 46417.

Mitchell, E., Spencer Chapman, M., Williams, N., Dawson, K.J., Mende, N., Calderbank, E.F., et al., 2022. Clonal dynamics of haematopoiesis across the human lifespan. Nature 606 (7913), 343–350.

Morley, J.E., Vellas, B., van Kan, G.A., Anker, S.D., Bauer, J.M., Bernabei, R., et al., 2013. Frailty consensus: a call to action. J. Am. Med. Dir. Assoc. 14 (6), 392–397.

Murad, M.H., Elamin, K.B., Abu Elnour, N.O., Elamin, M.B., Alkatib, A.A., Fatourechi, M. M., et al., 2011. The effect of vitamin d on falls: a systematic review and meta-analysis. J. Clin. Endocrinol. Metab. 96 (10), 2997–3006.

Nwadiugwu, M.C., 2020. Frailty and the risk of polypharmacy in the older person: enabling and preventative approaches. J. Aging Res. 2020, e6759521.

O'Caoimh, R., Sezgin, D., O'Donovan, M.R., Molloy, D.W., Clegg, A., Rockwood, K., et al., 2021. Prevalence of frailty in 62 countries across the world: a systematic review and meta-analysis of population-level studies. Age Ageing 50 (1), 96–104.

Ofori-Asenso, R., Chin, K.L., Mazidi, M., Zomer, E., Ilomaki, J., Zullo, A.R., et al., 2019. Global incidence of frailty and prefrailty among Community-dwelling older adults: a systematic review and Meta-analysis. JAMA Netw. Open 2 (8), e198398.

Panza, F., Seripa, D., Solfrizzi, V., Tortelli, R., Greco, A., Pilotto, A., et al., 2015. Targeting cognitive frailty: clinical and neurobiological roadmap for a single complex phenotype. J. Alzheimers Dis. 47 (4), 793–813.

Park, C., Mishra, R., Golledge, J., Najafi, B., 2021. Digital biomarkers of physical frailty and frailty phenotypes using Sensor-based physical activity and machine learning. Sensors 21 (16), 5289.

Patel, H.P., Clift, E., Lewis, L., Cooper, C., Patel, H.P., Clift, E., et al., 2017. Epidemiology of sarcopenia and frailty. Frailty and Sarcopenia - Onset, Development and Clinical Challenges. IntechOpen. ⟨https://www.intechopen.com/chapters/56116⟩.

Puts, M.T.E., Visser, M., Twisk, J.W.R., Deeg, D.J.H., Lips, P., 2005. Endocrine and inflammatory markers as predictors of frailty. Clin. Endocrinol. (Oxf.) 63 (4), 403–411.

Reid, N., Young, A., Hanjani, L.S., Hubbard, R.E., Gordon, E.H., 2022. Sex-specific interventions to prevent and manage frailty. Maturitas 164, 23–30.

Rockwood, K., Mitnitski, A., 2007. Frailty in relation to the accumulation of deficits. J. Gerontol. Ser. A 62 (7), 722–727.
Rockwood, K., Fox, R.A., Stolee, P., Robertson, D., Beattie, B.L., 1994. Frailty in elderly people: an evolving concept. CMAJ Can. Med. Assoc. J. 150 (4), 489–495.
Ross, A.C., Manson, J.E., Abrams, S.A., Aloia, J.F., Brannon, P.M., Clinton, S.K., et al., 2011. The 2011 report on dietary reference intakes for calcium and vitamin d from the institute of medicine: what clinicians need to know. J. Clin. Endocrinol. Metab. 96 (1), 53–58.
Shardell, M., Hicks, G.E., Miller, R.R., Kritchevsky, S., Andersen, D., Bandinelli, S., et al., 2009. Association of low vitamin D levels with the frailty syndrome in men and women. J. Gerontol. Ser. A. 64A (1), 69–75.
Spearman, C., 1904. The proof and measurement of association between two things. Am. J. Psychol. 15 (1), 72–101.
Spira, D., Buchmann, N., Nikolov, J., Demuth, I., Steinhagen-Thiessen, E., Eckardt, R., et al., 2015. Association of low lean mass with frailty and physical performance: a comparison between two operational definitions of Sarcopenia—Data from the Berlin aging study II (BASE-II). J. Gerontol. Ser. A. 70 (6), 779–784.
Spira, D., Buchmann, N., König, M., Rosada, A., Steinhagen-Thiessen, E., Demuth, I., et al., 2019. Sex-specific differences in the association of vitamin d with low lean mass and frailty: results from the Berlin aging study II. Nutrition 62, 1–6.
Strain, W.D., Down, S., Brown, P., Puttanna, A., Sinclair, A., 2021. Diabetes and frailty: an expert consensus statement on the management of older adults with type 2 diabetes. Diabetes Ther. 12 (5), 1227–1247.
Sze, S., Pellicori, P., Zhang, J., Weston, J., Clark, A.L., 2019. Identification of frailty in chronic heart failure. JACC Heart Fail 7 (4), 291–302.
Tao, X., Liu, L., Ma, P., Hu, J., Ming, Z., Dang, K., et al., 2024. Association of circulating very long-chain saturated fatty acids with cardiovascular mortality in NHANES 2003-2004, 2011-2012. J. Clin. Endocrinol. Metab. 109 (2), e633–e645.
Tenison, E., Henderson, E.J., 2020. Multimorbidity and frailty: tackling complexity in Parkinson's disease. J. Park Dis. 10 (s1), S85–S91.
Toepfer, S., König, M., Spira, D., Drewelies, J., Kreutz, R., Bolbrinker, J., et al., 2021. Sex differences in characteristics associated with potentially inappropriate medication use and associations with functional capacity in older participants of the Berlin aging study II. Gerontology 68 (6), 664–672.
Tomita, M.R., Mann, W.C., Fraas, L.F., Stanton, K.M., 2004. Predictors of the use of assistive devices that address physical impairments among community-based frail elders. J. Appl. Gerontol. 23 (2), 141–155.
Uchai, S., Andersen, L.F., Hopstock, L.A., Hjartåker, A., 2023. Body mass index, waist circumference and pre-frailty/frailty: the Tromsø study 1994−2016. BMJ Open 13 (2), e065707.
Uchmanowicz, I., Łoboz-Rudnicka, M., Szeląg, P., Jankowska-Polańska, B., Łoboz-Grudzień, K., 2014. Frailty in heart failure. Curr. Heart Fail Rep. 11 (3), 266–273.
Varrette, S., Cartiaux, H., Peter, S., Kieffer, E., Valette, T., Olloh, A., 2022. Management of an academic HPC & research computing facility: the ULHPC experience 2.0. Proceedings of the 2022 6th High Performance Computing and Cluster Technologies Conference [Internet]. Association for Computing Machinery, New York, NY, USA, pp. 14–24. https://doi.org/10.1145/3560442.3560445.
Vaz Fragoso, C.A., Enright, P.L., McAvay, G., Van Ness, P.H., Gill, T.M., FRAILTY, A.N.D., 2012. Respiratory impairment in older persons. Am. J. Med. 125 (1), 79–86.
Verschoor, C.P., Theou, O., Ma, J., Montgomery, P., Mossey, S., Nangia, P., et al., 2024. Age- and sex-specific associations of frailty with mortality and healthcare utilization in community-dwelling adults from Ontario, Canada. BMC Geriatr. 24 (1), 223.
Wang, C.W., Lebsack, A., Chau, S., Lai, J.C., 2019. The range and reproducibility of the liver frailty index. Liver Transpl. 25 (6), 841–847.
Wang, H. yu, Zhang, M., Sun, X., 2024. Sex-specific association between socioeconomic status, lifestyle, and the risk of frailty among the elderly in China. Front. Med. ⟨https://www.frontiersin.org/articles/10.3389/fmed.2021.775518⟩.
Welch, B.L., 1947. The generalization of 'student's' problem when several different population varlances are involved. Biometrika 34 (1–2), 28–35.
Woo, J., Yu, R., Wong, M., Yeung, F., Wong, M., Lum, C., 2015. Frailty screening in the community using the FRAIL scale. J. Am. Med Dir. Assoc. 16 (5), 412–419.
Xu, L., Zhang, J., Shen, S., Hong, X., Zeng, X., Yang, Y., et al., 2020. Association between body composition and frailty in elder inpatients. Clin. Interv. Aging 15, 313–320.
Xue, Q.L., 2011. The frailty syndrome: definition and natural history. Clin. Geriatr. Med. 27 (1), 1–15.
Yarnall, A.J., Sayer, A.A., Clegg, A., Rockwood, K., Parker, S., Hindle, J.V., 2017. New horizons in multimorbidity in older adults. Age Ageing 46 (6), 882–888.
Zhang, C., Bi, J., Soda, P., 2017. Feature selection and resampling in class imbalance learning: which comes first? An empirical study in the biological domain. International Conference on Bioinformatics and Biomedicine (BIBM) [Internet]. IEEE, pp. 933–938. ⟨https://ieeexplore.ieee.org/abstract/document/8217782⟩.
Zhou, J., Huang, P., Liu, P., Hao, Q., Chen, S., Dong, B., et al., 2016. Association of vitamin d deficiency and frailty: a systematic review and meta-analysis. Maturitas 94, 70–76.