# AI-Enhanced Tele-Rehabilitation: Predictive Modeling for Fall Risk and Treatment Efficacy in Balance Disorders

Efterpi Karapintzou [1], Vassilios Tsakanikas [1], Dimitrios Kikidis [2], Christos Nikitas [2], Brooke Nairn [3], Marousa Pavlou [4,5], Doris-Eva Bamiou[3,4], Themis Exarchos [6], Dimitrios I. Fotiadis[1,7,*], *Fellow, IEEE*

[1]Unit of Medical Technology and Intelligent Information Systems, Dept. of Materials Science and Engineering, University of Ioannina, Ioannina, Greece

[2] Dept. of Otorhinolaryngology, Head and Neck Surgery, National and Kapodistrian University of Athens, Hippocrateion General Hospital, Athens, Greece.

[3] Faculty of Brain Sciences, University College London (UCL) Ear Institute, University College London, London, United Kingdom

[4]Dept. of Neuro-Otology, University College London Hospitals, London, UK.

[5] Centre for Human and Applied Physiological Sciences, School of Basic and Medical Biosciences, Faculty of Life Sciences and Medicine, King's College London, Shepherd's House, Guy's Campus, London, United Kingdom

[6] Dept. of Informatics, Ionian University, Corfu, Greece.

[7] Biomedical Research Institute, Foundation for Research and Technology Hellas, FORTH, Ioannina, Greece

*Contact: fotiadis@uoi.gr, phone: +30 26510 09006

***Abstract*— This study focuses on the development of artificial intelligence models to enhance telerehabilitation practices. We utilized diverse datasets to create clinically relevant models for predicting two critical outcomes: fall risk and treatment effectiveness. By applying various machine learning techniques, including K-Nearest Neighbors (KNN), Random Forest, Decision Tree, Support Vector Machine (SVM), and XGBoost, our models demonstrated high accuracy, sensitivity, and specificity. Notably, the Random Forest model achieved an accuracy of 0.97 in predicting fall risk and 0.96 in assessing treatment effectiveness. These models equip clinicians with powerful tools for data-driven decision-making, ultimately improving patient outcomes in rehabilitation settings.**

***Keywords—AI prognostic analytics, Fall Risk, Treatment Effectiveness, Rehabilitation***

## I. INTRODUCTION

Balance, which is the ability to maintain the body's center of mass within its stability boundaries, is essential for movement and overall functional independence [1]. Balance disorders, often characterized by sensations of unsteadiness or vertigo, can result from a variety of causes, including specific medical conditions, certain medications, or dysfunctions within the inner ear or brain. These impairments can severely disrupt daily activities, leading to significant psychological and emotional challenges. There are over twelve recognized types of balance disorders, with some of the most common being positional vertigo, labyrinthitis, migrainous vertigo, vestibular neuritis, Ménière's disease, and Mal de Debarquement syndrome (MdDS) [2].

It is important to note that balance disorders and subsequent falls are also linked. Falls represent a major public health concern. Falls are the primary cause of both fatal and nonfatal injuries in older persons and are significant factors in illness and death, primarily resulting from decreased balance control. According to a survey conducted by the World Health Organization (WHO) [3], approximately 28-35% of those aged 65 and above experience falls annually, while the percentage increases to 32-42% for those aged 70 and above. Advancements in artificial intelligence (AI) offer novel approaches to balance assessment. These advancements have led to more precise diagnoses, personalized treatment methods, and improved monitoring and support systems.

This paper aims to advance the understanding and application of predictive modeling in tele-rehabilitation, specifically focusing on balance disorders. By applying state-of-the-art machine learning techniques and combining them with clinical data, the study presents a novel approach to assessing fall risk and evaluating treatment efficacy. The research contributes to the field by not only developing robust predictive models but also by validating these models in a real-world clinical context, thus bridging the gap between theoretical advancements and practical implementation. The findings presented in this paper hold significant implications for enhancing patient outcomes through personalized rehabilitation strategies and reducing healthcare costs associated with balance disorder management.

## II. STATE OF THE ART

Balance disorders pose a significant challenge in healthcare, particularly among elderly patients, where maintaining balance is crucial for daily activities and fall prevention. Recent technological advancements have introduced innovative approaches to assessing and managing balance disorders, shifting the focus from traditional methods to modern tools like wearable sensors and smartphone-based systems. These technologies offer new opportunities for more convenient and potentially more effective diagnosis and rehabilitation. However, it is essential to validate their reliability and effectiveness compared to established traditional methods before they can be widely adopted in clinical practice [4].

Recent advancements in the assessment of balance disorders have focused on integrating novel technologies and refining existing diagnostic methods. While traditional clinical assessments remain foundational, there has been a significant shift toward utilizing data-driven approaches, including artificial intelligence (AI) and machine learning, to enhance the accuracy and efficiency of balance evaluations. Despite these innovations, challenges such as data standardization, reliability of new tools compared to established methods, and the integration of multi-modal data remain prevalent. These hurdles must be addressed to fully realize the potential of these technologies in clinical practice, particularly for populations with diverse neurological conditions [5].

### A. Balance Assessment

Advancements in artificial intelligence (AI) offer novel approaches to balance assessment. Artificially intelligent methods are based on the evaluation of various balance control subsystems. One study has shown that AI can significantly improve the evaluation in balance performance but pointed out that further studies are needed to enhance the accuracy of assessment and find more applications, concluding that AI-based balance assessment systems have a potentially huge impact on clinical practice, rehabilitation and fall risk monitoring. [6].

Understanding the relationship between strength, balance, and functional recovery after a stroke is crucial for effective rehabilitation. In a recent study, the Five Times Sit-to-Stand (FTSTS) test, a key measure of functional ability post-stroke, was used to evaluate the influence of lower-limb strength and balance. The findings revealed that both knee extensor strength and balance metrics independently predicted FTSTS performance, highlighting the importance of addressing both factors in post-stroke rehabilitation programs.[7].

The use of valid and reliable measures is essential for assessing balance in the elderly, and the modified Clinical Test of Sensory Interaction in Balance (mCTSIB) provides valuable insights in this regard. Despite moderate reliability, the mCTSIB demonstrates significant potential as a quick and independent tool for initial postural evaluation. It offers practical advantages over traditional clinical tests, making it a promising option for assessing balance in older adults.[1]. Other researchers have also proposed new frameworks combining IoT, AI algorithms, and big data analytics in an effort to guarantee the safety and wellbeing of elderly individuals, including fall monitoring and emotion recognition [8].

Technological advancements have expanded into home-based rehabilitation, offering systems that integrate sensors and deep learning to provide objective and convenient methods for balance evaluation. These systems utilize deep learning models to accurately estimate a subject's Center of Mass (CoM) position using data from depth cameras. Portable and cost-effective, these technologies enable on-demand balance assessments both at home and in clinical settings. This approach has the potential to significantly reduce the frequency of in-person visits, thereby lowering costs for patients and healthcare providers [9].

Moreover, the accuracy of the machine learning algorithms to detect a fall is impressive, addressing concerns about false alarms. Chelli and Pätzold [10] achieved a perfect accuracy of 100% with zero false alarms using quadratic Support Vector Machine (SVM) and ensemble bagged tree (EBT) on acceleration and angular velocity data from two publicly available datasets.

The timely identification of fall events is crucial for the effectiveness and safety of fall prevention systems, as it allows the control device to promptly respond and prevent the occurrence of falls. Neural network predictive models provide insights into falls due to unexpected perturbations during regular walking [11]. The study simulates falls, recording 3D skeletal data and calculating the average time from a disturbance to the initiation of a fall. This is very relevant for earlier intervention strategies in falling prevention efforts among elderly individuals with neurological disorders.

### B. Rehabilitation

Rehabilitation aims to optimize functional capacity and minimize disability for individuals with health conditions, considering their unique circumstances and environment. The integration of remote monitoring through wearable multimodal devices and machine learning enables objective tracking of clinical progress, enhancing the precision and personalization of rehabilitation interventions [12]. The emphasis on explainable and interpretable AI further supports clinical decision-making by providing insights into the factors driving functional recovery predictions, particularly in areas like upper limb recovery post-stroke. The COVID-19 pandemic has underscored the importance of unsupervised home-based rehabilitation, driving the development of innovative systems that prioritize individualization, patient engagement, and adherence to ethical guidelines [13].

Rehabilitation programs are fundamental for the enhancement of quality in life of individuals with balance disorders. The major two reasons, such as limited resources for clinical setup and social isolation, have made rehabilitation mainly carried out at residences, which have often led to poor execution and dropping out of exercises without proper motivation and direction. There were seven categories of balance rehabilitation interventions described by Saraiva *et al.* [14]: conventional balance exercises, gym-based interventions, vibration therapy, rhythmic auditory stimulation training, boxing therapy, dual-task training, and technology-based interventions. Since all these approaches often work in their individual way, the recovery of balance in stroke survivors can be greatly affected, thus also enhancing their general wellbeing and functional capacity.

Examining the efficacy of home-based exercise programs (HEPs) incorporating augmented reality (AR), it was found [15] that stroke patients receiving this program showed significantly higher improvements in balance and reduction in fear of falling compared to controls. A virtual coaching system [16] devised to manage balance disorders integrates sensing devices and augmented reality technologies that deliver automated personal feedback during exercises. Being able to achieve high accuracy in the assessment of posture and gait metrics, this could possibly revolutionize home rehabilitation of balance disorders. Furthermore, a data-driven scoring methodology, based on machine-learning techniques, designed to increase the accuracy in exercise assessment, using lessons learned from expert-monitored sessions, enhances the consequent reliability of the scoring module of the HOLOBALANCE system [17].

The assessment and rehabilitation of balance disorders are evolving through the integration of advanced technologies and human-centered personalized practices. Wearable sensors, artificial intelligence, and virtual coaching are paving the way for innovative, individualized interventions aimed at improving balance, and ultimately, enhancing wellbeing and quality of life. Although significant progress has been made, ongoing research and validation are essential to ensure the effectiveness and safety of these interventions, particularly in more diverse populations.

## III. Materials

### A. Datasets

Four datasets were collected from participants experiencing balance disorders, with two of these datasets generated by the National and Kapodistrian University of Athens (NKUA). The primary objective of the NKUA datasets is to examine factors that may influence rehabilitation outcomes, as measured by changes in the Functional Gait Assessment (FGA) and Dizziness Handicap Inventory (DHI) scores over an 8-week period. While previous research has not explicitly identified these factors, our hypothesis suggests that they may impact fall risk and treatment effectiveness. Specifically, these datasets evaluate two independent populations undergoing individualized 8-week vestibular rehabilitation programs, with each dataset offering insights into the relationship between fall risk and various health characteristics in individuals with Mild Cognitive Impairment (MCI) and other vestibular disorders, as detailed in the following subsection.

The third dataset was collected by University College London (UCL), providing demographic and clinical information along with pre- and post-physiotherapy scores for the Dizziness Handicap Inventory (DHI), Visual Analog Scale (VAS), and Functional Gait Assessment (FGA). Finally, the fourth dataset was generated as part of the HOLOBALANCE project [MISSING REFERENCE], which aimed to develop and validate an augmented-reality virtual coaching platform for personalized rehabilitation monitoring in the aging population with balance disorders. Overall, these datasets cover a wide range of approaches, strengthening the validity of the prediction models created to evaluate the likelihood of treatment efficacy and fall risk. This multicentre collaboration underscores the significance of interdisciplinary research in enhancing telerehabilitation practices and improving patient outcomes.

### B. Population

NKUA provided datasets comprising 248 subjects. This dataset includes detailed demographic information, symptom severity, clinical signs, balance and fall assessments, as well as associated co-morbidities.

The first dataset (NKUA V1), comprising 104 subjects, includes patients suffering from various vestibular disorders. This includes 44 patients with Persistent Postural-Perceptual Dizziness (PPPD), 20 with Vestibular Migraine (VM), 19 with Meniere's disease, 3 with Acoustic Neuroma (AN), 7 with Benign Paroxysmal Positional Vertigo (BPPV), 5 with Bilateral Vestibular Weakness (BVW), 2 with Central Vestibular Disorder (CVD), 1 with Mal de Debarquement syndrome (MDDS), 1 with Multiple Sclerosis (MS), and 2 with Mild Traumatic Brain Injury (MTBI).

The second dataset (NKUA V2) comprises 144 subjects, with 74 diagnosed with PPPD, 26 with VM, 20 with BPPV, 8 with Meniere's disease, 7 with Neuritis, 4 with BBPV, 3 classified as Fallers, 1 with CVD, and 1 with AN. These data are crucial for our study in order to define associations between fall risk and various factors within the context of MCI and vestibular disturbances.

The dataset provided by UCL consists of a total of 93 subjects with VM, with or without Traumatic Brain Injury (TBI).

The HOLOBALANCE dataset includes data from 129 subjects at total: 65 for the intervention group and 64 for the control group. All participants are over 40 years old. This dataset contains comprehensive clinical evaluation data and sensor data, demographic information as well as a variety of clinical exams and questionnaires: EQ-5D5L for measuring health-related quality of life, MoCA for assessing cognitive function, SUS for evaluating system usability, FGA for analyzing gait, ABC for measuring balance confidence, RAPA for assessing physical activity, and Mini-Best for evaluating balance.

Table I provides population demographics, mentioning the total number of individuals and their respective age groups. Collectively, these datasets establish a strong foundation, facilitating a thorough examination of the factors that influence the efficacy of treatment and the risk of falling in patients with balance disorders.

*Table I. Population Statistics.*

| | NKUA V1 | NKUA V2 | UCL | HOLOB |
|---|---|---|---|---|
| **Number of patients** | 104 | 144 | 93 | 129 |
| **Mean Age** | 50 years | 58 years | 49 years | 69 years |
| **Min Age** | 20 years | 20 years | 16 years | 43 years |
| **Max Age** | 83 years | 88 years | 89 years | 43years |

## IV. Methodology

### A. Machine Learning Pipeline

The main structure of our study's artificial intelligence system consists of the following components: (i) data loading and pre-processing, (ii) building of classifiers (model training), (iii) performance evaluation, and (v) explainable AI using SHapley Additive exPlanations (SHAP) values. The machine learning pipeline is also illustrated in Fig. 1.

This machine learning pipeline is initiated by loading and preprocessing the dataset, which includes encoding categorical features, detecting feature types, conducting statistical tests for feature selection, handling missing values, and scaling numerical features. The 10-fold cross-validation process was implemented by partitioning the dataset, allocating 70% of the data for training and 30% for testing. This methodology guarantees that the model's effectiveness is assessed on a substantial and varied dataset. This approach ensured a thorough evaluation of the model's performance across all classes, with statistical certainty.

The pipeline deals with imbalance issues by utilizing the Synthetic Minority Oversampling Technique (SMOTE). This approach ensures a more balanced representation of the data, enhancing the reliability of the analysis and the accuracy of the results.

After the data preparation phase is completed, the pipeline proceeds to the training of numerous classifiers. The classifiers include Logistic Regression, Support Vector Machines, K-Nearest Neighbors (KNN), Decision Tree, Random Forest, Naïve Bayes, Gradient Boosting, XG Boost and Neural Networks. GridSearchCV is utilized for each model to find the best parameter setting of the model through hyperparameter optimization. Afterward, evaluation of the models took place to get significant parameters such as specificity, sensitivity, and accuracy. The models with outstanding results are preserved and retained for future use.

Finally, the pipeline is implementing SHAP to explain the model predictions. SHAP method is a widely used tool nowadays for most researchers to provide a measure of the contribution of each feature towards a classification outcome. This step enables us to understand the model's behavior at a deeper level, making the model more interpretable.

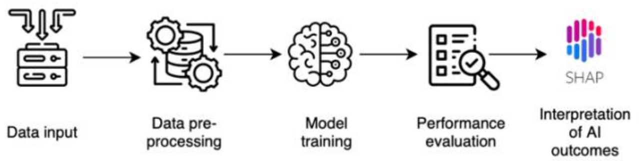


*Fig. 1. Machine Learning Pipeline.*

## B. *Clinical Endpoints*

We utilized the datasets to construct predictive models targeting two main endpoints: the probability of treatment effectiveness and the probability of risk of falls.

Initially, our main objective was to predict the variation of the FGA scale between the pre-intervention and post-intervention stages to evaluate the likelihood of falls. This entailed examining the FGA scores before and after the intervention to ascertain the effect of the interventions on patients' gait and balance, thereby approximating their susceptibility to falling.

Our objective was to predict the efficacy of treatment by utilizing the MoCA questionnaire. Through the assessment of cognitive function at the beginning of the rehabilitation program, our objective was to determine the impact of initial cognitive abilities on the total effectiveness of the treatment.

Moreover, we constructed models to predict the difference in the DHI scale before and after the intervention, with the specific aim of assessing treatment efficacy. This method enabled us to assess enhancements in patients' subjective impairment caused by dizziness, offering a quantifiable indication of the intervention's influence on their day-to-day activities and overall well-being.

Finally, our objective was to forecast the disparity in the EQ-5D scale between the pre and post-intervention stages at the baseline, in order to assess the efficacy of the treatment. The EQ-5D scale quantifies the health-related quality of life, allowing us to evaluate the impact of interventions on patients' overall well-being and health status by analysing deviations from the initial state.

*Table II* presents an overview of these endpoints, including the specific models used, the datasets employed, and the attained metrics.

*Table II. Overall clinical endpoints.*

| Group Category | Clinical Tools (Target) | Justification |
|---|---|---|
| **Risk of Fall** | DIF_FGA (Difference-change of FGA scale before and after intervention) | These tools assess balance, gait functionality, mobility, and confidence, which are crucial in evaluating the risk of falls. |
| **Treatment Effectiveness** | MoCA (MoCA scale at the baseline),<br>DIF_DHI (Difference-change of DHI scale before and after intervention)<br>DIF_5Q5D | These measures provide insights into cognitive function, disability level, self-care capability, motivational factors, Pain and anxiety levels and dizziness level which are key indicators of overall treatment effectiveness. |

# V. Results

This section provides a comprehensive summary of the predictive models developed and reported in the previous sections. More specifically, Table III presents an outline of the models, along with the relative information about the utilized dataset and relative metrics. Certain models are identified as being particularly pertinent for clinical relevance.

The generated models were highly effective in predicting the probability of falls. The application of Gradient Boosting on the NKUA V2 dataset yielded solid results, with an accuracy of $0.97 \pm 0.03$, sensitivity of $0.95 \pm 0.05$, and specificity of $0.95 \pm 0.04$. In this regard, a very important feature was identified: Functional Gait Assessment at baseline. Additionally, the K-Nearest Neighbors (KNN) classifier employing the NKUA V1 dataset, attained an accuracy of $0.85 \pm 0.12$, a sensitivity of $0.82 \pm 0.18$, and a specificity of $0.79 \pm 0.22$, highlighting the Comorbidities and the FGA at the baseline as key features.

For the assessment of treatment effectiveness, multiple models were trained to assess cognitive and physical health metrics. A Random Forest classifier using the NKUA V2 dataset derived for DIF_DHI scores gave an accuracy of $0.96 \pm 0.04$ with a sensitivity of $0.95 \pm 0.05$ and specificity of $0.90 \pm 0.10$. Similarly, this model highlighted education and Dizziness Handicap Inventory as key predictors. The Random Forest model for the NKUA V1 dataset reached an accuracy of $0.96 \pm 0.04$, a sensitivity of $0.92 \pm 0.08$, and a specificity of $0.88 \pm 0.12$, where the most important features were Comorbidities and DHI.

The overall accuracy of the models developed in this project are very high for different endpoints. Models to be highlighted for use in clinical practice will be very important in enhancing clinical decision-making and improvement of patient outcomes in rehabilitation settings. Thus, these models will prove to be useful in providing personalized patient care, allowing therapies to be customized based on individual risk profiling. Incorporating these models into clinical processes has the potential to enhance therapies and improve the management of balance problems.

*Table III.Summary of the best classifiers and their important features.*

| **Endpoint** | **Source** | **Target** | **Classifier** | **Accuracy** | **Sensitivity** | **Specificity** | **Feature (Baseline)** |
|---|---|---|---|---|---|---|---|
| **Risk of fall** | NKUA (V1) | DIF_FGA | KNN | 0.85 ± 0.12 | 0.82 ± 0.18 | 0.79± 0.22 | Comorbidities & FGA |
| | **NKUA (V2)** | **DIF_FGA** | **GB** | **0.97 ± 0.03** | **0.95 ± 0.05** | **0.95 ± 0.04** | **FGA** |
| **Treatment Effectiveness** | HOLOBALANCE | DIF_MoCA | KNN | 0.71 ± 0.13 | 0.78 ± 0.16 | 0.62 ± 0.20 | MoCA |
| | NKUA (V1) | DIF_DHI | RF | 0.96 ± 0.04 | 0.92 ± 0.08 | 0.88 ± 0.12 | Comorbidities, DHI |
| | **NKUA (V2)** | **DIF_DHI** | **RF** | **0.96 ± 0.04** | **0.95 ± 0.05** | **0.90 ± 0.10** | **Education & DHI** |
| | HOLOBALANCE | DIF_5Q5D | GB | 0.76 ± 0.18 | 0.82 ± 0.15 | 0.71 ± 0.29 | Falls past year |
| | UCL (VM) | DIF_DHI | DT | 0.93 ± 0.07 | 0.91± 0.04 | 0.95± 0.05 | Fall |

## A. Risk of Fall

The SHAP values for each feature in the NKUA V1 and V2 datasets provide a thorough comprehension of the overall influence of input variables on the model's predictions. By combining SHAP values for each feature, we obtain insights into the comparative significance of various qualities in impacting the decision-making process of the model. This thorough examination reveals the characteristics that have the most significant impact on the model's outputs and provides valuable insights between the input features and the prediction. For the NKUA V1, the **Comorbidities and FGA at baseline** are the most significant features. Whereas, the **FGA at baseline** feature is the most significant aspect in NKUA V2, as seen by the SHAP values displayed in Fig. 2 and Fig. 3.

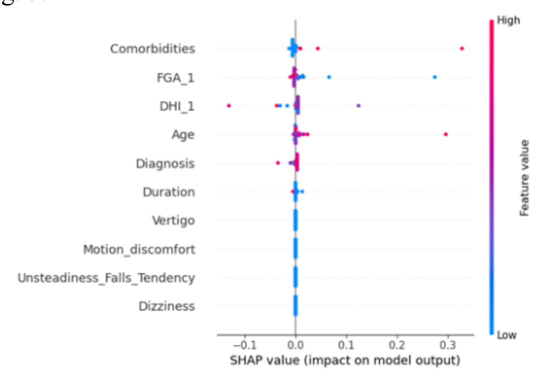


*Fig. 2. Risk of fall endpoint: SHAP value of features in the NKUA(V1) dataset.*

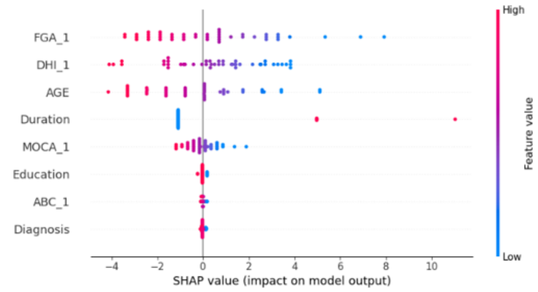


*Fig. 3.Risk of fall endpoint: SHAP value of features in the NKUA(V2) dataset.*

Table IV summarizes the important features identified from each model, outlining both the distinct features from each version of the dataset and the common features that emerge as critical across both models. The identification of common important features is key, for they can afford a robust base with which to understand what factors influence risk of falls.

*Table IV. The most important features of the Risk of Fall endpoint.*

| **NKUA (V1)- DIF_FGA** | **NKUA(V2) – DIF_FGA** |
|---|---|
| Comorbidities | FGA |
| FGA | DHI |
| DHI | Age |
| Age | Duration |
| Diagnosis | MOCA |
| Duration | Education |
| Vertigo | ABC |
| Motion Discomfort | Diagnosis |
| Unsteadiness_Falls Tendency | |
| Dizziness | |

Both models identify the FGA and DHI at baseline as critical features, emphasizing their significance in predicting the risk of fall endpoint. This association highlights the need of incorporating these assessments into practice and falling prevention programs. More specifically, the NKUA V1 model emphasizes a wider range of clinical and symptomatic data, including comorbidities, diagnosis, vertigo, movement discomfort, instability, tendency to fall and dizziness. This model's comprehensive approach suggests a detailed consideration of various clinical signs and symptoms to assess fall risk accurately. Conversely, the NKUA V2 model incorporates age, duration of symptoms, the Montreal Cognitive Assessment (MOCA), education, and the Activities-specific Balance Confidence (ABC) scale as significant predictors. These factors suggest that cognitive function, educational background, and perceived balance confidence play a very important role in the risk of falling for each individual. Moreover, age and duration of symptoms support demographic and clinical characteristics in assessing fall risk.

This comparison and the identification of common important features for a future overall model is what highlights the various elements that contribute to and/or influence the risk of fall. By combining clinical and cognitive testing, predictive models become more powerful, providing physicians with vital tools for preventing and managing falls in rehabilitation.

### B. Treatment Effectiveness

Correspondingly SHAP values for all features in the treatment effectiveness endpoint datasets provide a comprehensive understanding of the overall impact of the input variables on the predictions of this model. This detailed insight allows us to quantify the contribution of each attribute, thereby enhancing the interpretability and transparency of the model's decision-making process.

The most crucial factors in the NKUA V1 scenario are the comorbidities and DHI at baseline characteristics. The Education and DHI at baseline feature are the most prominent aspect in NKUA V2, as indicated by the SHAP values presented in Fig. 4 and Fig. 5.

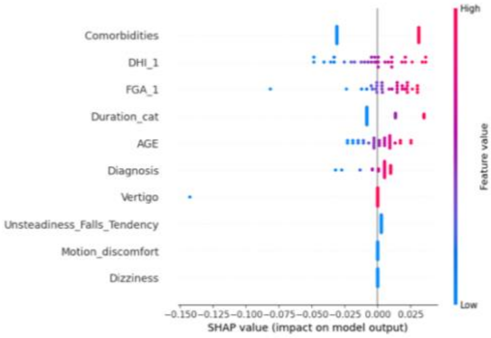


*Fig. 4.Treatment effectiveness endpoint: SHAP value of features in the NKUA(V1) dataset.*

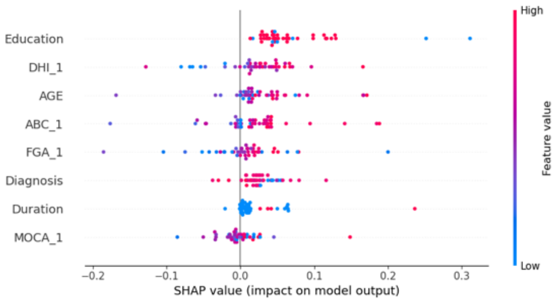


*Fig. 5.Treatment effectiveness endpoint: SHAP value of features in the NKUA(V2) dataset.*

In relation to the UCL (VM) dataset, the Fall feature holds the utmost importance in this scenario as illustrated in Fig. 6. This underscores the critical role that this feature plays in accurately predicting and understanding the risk of falls within this dataset.

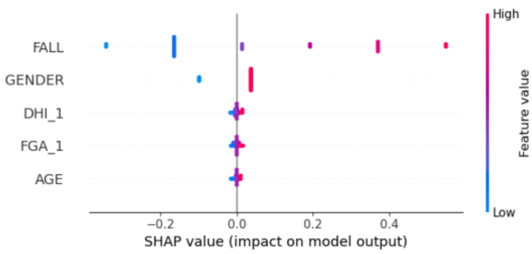


*Fig. 6.Treatment effectiveness endpoint: SHAP value of features in the UCL (VM) dataset.*

The HOLOBALANCE dataset produced two different targets in the treatment efficacy endpoint: the difference in MoCA scores and the difference in 5Q5D scores. Regarding the first analysis according to the results seen in Fig. 7a the most relevant feature is MoCA total. Specifically, this refers to the total MoCA scale score at baseline. On the other hand, Fig. 7b, which reflects on DIF_5Q5D, it demonstrates that the number of falls that occurred in the past year is the most important characteristic.

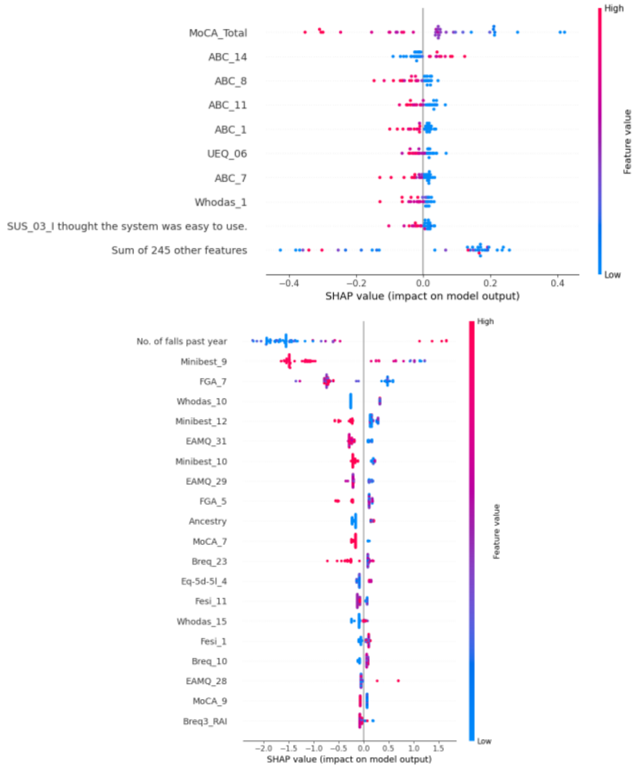


*Fig. 7. Treatment effectiveness endpoint: SHAP value of features in the HOΛOBAΛANΨE dataset (a.DIF_MoCA, b.DIF_5Q5D).*

The most important features of the treatment effectiveness endpoint were analyzed and compared. This comparison reveals important derived from the models developed using the NKUA V1, NKUA V2, UCL (VM), and HOLOBALANCE datasets. In fact, Table V outlines the key features identified from each model.

*Table V.Most important features of Treatment effectiveness endpoint.*

| NKUA(V1)-DIF_DHI | NKUA (V2)-DIF_DHI | UCL (VM) - DIF_DHI | HOLOB.-DIF_MoCA | HOLOB.-DIF_5Q5D |
|---|---|---|---|---|
| Comorbidities | Education | FALL | MoCA_Total | No. of falls past year |
| DHI | DHI | Gender | ABC_14 | Minibest_9 |
| FGA | Age | DHI | ABC_8 | FGA_7 |
| Duration | ABC | FGA | ABC_11 | Whodas_10 |
| Age | FGA | Age | ABC_1 | Minibest_12 |
| Diagnosis | Diagnosis | | UEQ_06 | EAMQ_31 |
| Vertigo | Duration | | ABC_7 | Minibest_10 |
| Unsteadiness, Falls Tendency | MOCA | | Whodas_1 | EAMQ 29 |
| Motion Discomfort | | | SUS_03_I thought the system was easy to use | FGA 5 |
| Dizziness | | | UEQ_17 | Ancestry |

The FGA feature is present in almost all the models, pointing to their importance in relation to the treatment effectiveness endpoint. Moreover, another very important feature in this endpoint is the DHI, since this feature is identified as critical by the NKUA V1, NKUA V2, and UCL VM models.

In particular, the NKUA V1 model emphasizes a wide range of demographic and clinical signs, including comorbidities, duration of symptoms, age, diagnosis, vertigo, tendency to fall, discomfort from movement, and dizziness. While the NKUA model V2 is on the same wavelength. Specifically, it also incorporates demographic and clinical factors such as education, age, ABC, diagnosis, duration and MoCA. These models incorporating each participant's demographic and clinical cues suggest a more holistic approach to evaluating treatment outcomes in this scenario. The UCL (VM) model uniquely outlines fall risk, gender and age, alongside standard clinical assessments like DHI and FGA as important features that play a key role in treatment effectiveness.

The HOLOBALANCE - DIF_MoCA model integrates cognitive, balance, and user experience metrics, including MoCA_Total, various ABC measures, the User Experience Questionnaire (UEQ), and system usability. The focus of this model on the cognitive and user experience elements provides a comprehensive view of the effectiveness of the treatment. The HOLOBALANCE – DIF_5Q5D model, on the other hand, points out such factors as the number of falls the last year, various items of MiniBest and FGA scales, WHO Disability Assessment Schedule, Whodas, EuroQoL, EAMQ, and ancestry. This model's comprehensive approach to evaluating treatment outcomes underscores the importance of integrating multiple dimensions of patient data, including physical, cognitive, and demographic information.

Comparative observations highlight treatment efficacy factors and suggest tailored interventions. Clinical and cognitive assessments combine with user experience for robust prediction models, optimizing rehabilitation procedures for optimal patient outcomes.

## VI. Conclusions

The developed models specifically target important clinical outcomes, such as the risk of falls and treatment effectiveness. Moreover, they have the potential to improve the clinical decision-making process and patient rehabilitation outcomes.

For each endpoint, we have developed clinically relevant models that predict future outcomes using baseline features. For instance, within the NKUA dataset, DIF_FGA and DIF_DHI assessed the difference-change of each scale at the beginning and end of the 8-weeks. The UCL dataset used DIF_DHI to measure changes at the beginning and after an average of 33 weeks. The HOLOBALANCE dataset employed DIF_MoCA and DIF_5Q5D to assess scales at baseline (week 3) and at the end of the period (week 9).

For risk of falls, the developed models demonstrate strong robustness in terms of high accuracy, sensitivity, and specificity. The model that utilized the Random Forest technique, attained an accuracy of $0.97 \pm 0.03$ from the NKUA V2 dataset. These models leverage important variables such as FGA and DHI scales for predicting and mitigating fall risks in patients undergoing rehabilitation.

Additionally, our models demonstrate high accuracy and reliability for the treatment effectiveness endpoints. In particular, the Random Forest model targeting DIF_DHI from the datasets achieved impressive results with an accuracy of $0.96 \pm 0.04$, highlighting the importance of features such as comorbidities, DHI and education. Indeed, the most important features for the treatment effectiveness endpoint highlighted by the model set were the FGA and DHI scales.

In conclusion, our work has made substantial progress in detecting the key features that influence clinical outcomes through comprehensive model comparisons. The models developed in this study offer practical and effective tools for rehabilitation, enhancing clinical decision-making and ultimately improving patient outcomes.

## VII. Discussion

### SHAP values and their clinical relevance

The use of SHAP values across our models has provided crucial insights into the influence of specific features on model predictions.

For the **risk of fall** endpoint, the SHAP values highlight the Functional Gait Assessment (FGA) and the Dizziness Handicap Inventory (DHI) as the most significant features across the NKUA V1 and V2 datasets. The FGA at baseline consistently emerged as a crucial predictor in both models, underscoring its importance in assessing balance and gait functionality. Additionally, comorbidities, diagnosis, and symptom duration were significant in the NKUA V1 dataset, suggesting a comprehensive evaluation of a patient's medical background is vital for accurate fall risk prediction. The NKUA V2 dataset emphasized cognitive assessments such as the Montreal Cognitive Assessment (MOCA) and the Activities-specific Balance Confidence (ABC) scale, indicating that cognitive factors also play a significant role in fall risk.

For the **treatment effectiveness** endpoint, the SHAP analysis identified comorbidities, education level, and baseline DHI scores as pivotal in predicting outcomes. The consistent significance of the DHI across different datasets (NKUA V1, V2, and UCL) highlights its reliability in evaluating the impact of dizziness on daily activities and its utility in tracking treatment progress. The inclusion of educational background as a significant feature in the NKUA V2 dataset suggests that socioeconomic factors might influence treatment adherence and effectiveness. The Holobalance datasets added another layer by incorporating user experience metrics such as system usability (SUS) and user experience (UEQ), which are critical in tailoring and optimizing tele-rehabilitation programs.

### Similarities Among Datasets and Endpoints

Analyzing the similarities among the datasets and endpoints reveals several key insights:

1. **Consistency of Critical Features**: The FGA and DHI were recurrently significant across different datasets for both endpoints. This consistency reinforces their reliability and importance in clinical assessments related to balance disorders.

2. **Integration of Cognitive and Physical Assessments**: The inclusion of cognitive measures (e.g., MOCA) and balance confidence (e.g., ABC) across multiple models indicates the intertwined nature of cognitive and physical health in rehabilitation outcomes. This integration is essential for comprehensive patient evaluations and personalized rehabilitation strategies.
3. **Holistic Patient Profiles**: Both endpoints benefited from a holistic approach that considered demographic, clinical, and cognitive factors. This comprehensive profiling ensures that predictive models are robust and applicable to diverse patient populations, enhancing their generalizability and clinical utility.

**Clinical Relevance**

The SHAP values provide clinically relevant information that can directly impact patient care:

- **Personalized Treatment Plans**: By understanding which features most significantly influence outcomes, clinicians can tailor rehabilitation programs to address specific patient needs, improving adherence and effectiveness.
- **Risk Stratification**: Identifying high-risk patients through significant predictors like FGA and DHI allows for early interventions, potentially preventing falls and associated complications.
- **Enhanced Monitoring**: Continuous assessment using validated scales like FGA and DHI, coupled with cognitive and user experience evaluations, provides a multidimensional view of patient progress, enabling timely adjustments to treatment plans.
- **Data-Driven Decision Making**: The insights gained from SHAP values enhance the transparency and interpretability of AI models, fostering greater trust among clinicians in utilizing these tools for decision-making.

## VIII. Future Work

Future work will include incorporating additional retrospective datasets to improve and validate the current models. These datasets will be essential to improve the complexity and accuracy of our predictive models, guaranteeing their power and practicality in clinical settings. In essence, our goal is to improve the predictive capabilities of our existing models. This approach will significantly improve the efficiency of our project and therefore the patient outcomes, giving more reliable tools to the clinicians.


## Acknowledgment

This work has received funding from the European Union's Horizon 2021 research and innovation actions under grant agreement No 101057747, as part of the TeleRehaB DSS project.


## References


[1] E. Antoniadou *et al.*, 'Reliability and validity of the mCTSIB dynamic platform test to assess balance in a population of older women living in the community', *J. Musculoskelet. Neuronal Interact.*, vol. 20, no. 2, pp. 185–193, 2020.

[2] 'Balance Disorders — Causes, Types & Treatment | NIDCD'. Accessed: Mar. 12, 2024. [Online]. Available: https://www.nidcd.nih.gov/health/balance-disorders

[3] 'WHO global report on falls prevention in older age'. Accessed: Jul. 29, 2024. [Online]. Available: https://www.who.int/publications/i/item/9789241563536

[4] A. Gawronska, A. Pajor, E. Zamyslowska-Szmytke, O. Rosiak, and M. Jozefowicz-Korczynska, 'Usefulness of Mobile Devices in the Diagnosis and Rehabilitation of Patients with Dizziness and Balance Disorders: A State of the Art Review', *Clin. Interv. Aging*, vol. 15, pp. 2397–2406, Dec. 2020, doi: 10.2147/CIA.S289861.

[5] A. Zampogna *et al.*, 'Fifteen Years of Wireless Sensors for Balance Assessment in Neurological Disorders', *Sensors*, vol. 20, no. 11, Art. no. 11, Jan. 2020, doi: 10.3390/s20113247.

[6] 'Assessment of Balance Control Subsystems by Artificial Intelligence | IEEE Journals & Magazine | IEEE Xplore'. Accessed: Mar. 26, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8960402

[7] B. F. Mentiplay, R. A. Clark, K. J. Bower, G. Williams, and Y.-H. Pua, 'Five times sit-to-stand following stroke: Relationship with strength and balance', *Gait Posture*, vol. 78, pp. 35–39, May 2020, doi: 10.1016/j.gaitpost.2020.03.005.

[8] M. J. Al Nahian, T. Ghosh, M. N. Uddin, Md. M. Islam, M. Mahmud, and M. S. Kaiser, 'Towards Artificial Intelligence Driven Emotion Aware Fall Monitoring Framework Suitable for Elderly People with Neurological Disorder', in *Brain Informatics*, M. Mahmud, S. Vassanelli, M. S. Kaiser, and N. Zhong, Eds., Cham: Springer International Publishing, 2020, pp. 275–286. doi: 10.1007/978-3-030-59277-6_25.

[9] W. Wei, C. Mcelroy, and S. Dey, 'Using Sensors and Deep Learning to Enable On-Demand Balance Evaluation for Effective Physical Therapy', *IEEE Access*, vol. 8, pp. 99889–99899, 2020, doi: 10.1109/ACCESS.2020.2997341.

[10] A. Chelli and M. Pätzold, 'A Machine Learning Approach for Fall Detection and Daily Living Activity Recognition', *IEEE Access*, vol. 7, pp. 38670–38687, 2019, doi: 10.1109/ACCESS.2019.2906693.

[11] R. Mori, T. Furukawa, Y. Makino, and H. Shinoda, 'Measurement of Disturbance-Induced Fall Behavior and Prediction Using Neural Network', in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Jul. 2020, pp. 2142–2149. doi: 10.1109/SMC42975.2020.9283245.

[12] H. De Cannière *et al.*, 'Wearable Monitoring and Interpretable Machine Learning Can Objectively Track Progression in Patients during Cardiac Rehabilitation', *Sensors*, vol. 20, no. 12, Art. no. 12, Jan. 2020, doi: 10.3390/s20123601.

[13] I. Vourganas, V. Stankovic, and L. Stankovic, 'Individualised Responsible Artificial Intelligence for Home-Based Rehabilitation', *Sensors*, vol. 21, no. 1, Art. no. 1, Jan. 2021, doi: 10.3390/s21010002.

[14] J. Saraiva, G. Rosa, S. Fernandes, and J. B. Fernandes, 'Current Trends in Balance Rehabilitation for Stroke Survivors: A Scoping Review of Experimental Studies', *Int. J. Environ. Res. Public. Health*, vol. 20, no. 19, Art. no. 19, Jan. 2023, doi: 10.3390/ijerph20196829.

[15] 'Effects of the home-based exercise program with an augmented reality system on balance in patients with stroke: a randomized controlled trial: Disability and Rehabilitation: Vol 45, No 10'. Accessed: Mar. 26, 2024. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/09638288.2022.2074154

[16] V. D. Tsakanikas *et al.*, 'Evaluating the Performance of Balance Physiotherapy Exercises Using a Sensory Platform: The Basis for a Persuasive Balance Rehabilitation Virtual Coaching System', *Front. Digit. Health*, vol. 2, Nov. 2020, doi: 10.3389/fdgth.2020.545885.

[17] V. Tsakanikas *et al.*, 'Automated Assessment of Balance Rehabilitation Exercises With a Data-Driven Scoring Model: Algorithm Development and Validation Study', *JMIR Rehabil. Assist. Technol.*, vol. 9, no. 3, p. e37229, Aug. 2022, doi: 10.2196/37229.