On Gait-Based Identification of Persons During Winter Conditions

Grigorij Rego, Yulia Vahroeva
Petrozavodsk State University
Petrozavodsk, Russia
regoGr@yandex.ru, vakhroeva@list.ru

Alexei Falev
Petrozavodsk State University
Petrozavodsk, Russia
alexey000035@gmail.com

***Abstract*—This paper investigates the effectiveness of gait-based identification across different seasons which is especially important for the northern territories. The study highlights the uniqueness of gait patterns, influenced by anatomical and physiological characteristics, and their potential for non-invasive identification methods. It addresses the critical question of whether a model trained on summer data can identify individuals in winter, revealing difficulties due to seasonal changes affecting gait patterns. The research employs a comprehensive dataset collected using smartphones equipped with accelerometers and gyroscopes, capturing various gait parameters across different terrains and seasons. The paper explores data filtering techniques and machine learning algorithms, including decision trees, k-nearest neighbors, and random forests, to analyze gait data. The results demonstrate the impact of seasonal variations on model performance, underscoring the need for diverse training datasets to enhance identification accuracy. The study's findings suggest that while gait identification holds promise for various applications, its reliability is contingent on the inclusion of diverse seasonal data in model training. The study contributes to the understanding of the potential and limitations of gait analysis and contributes to the development of digital assistants for tracking human motor activity.**

I. Introduction

Human identification is the process of determining an individual's identity [1]. Gait-based identification, a biometric technology that identifies individuals by their walking patterns, has emerged as a significant area of research in the field of security and surveillance. This technology is predicated on the principle that each person has a unique way of walking, which can be captured and analyzed to identify them. Gait identification technology is founded on a premise that an individual's walking pattern is unique and can be quantified through various parameters such as stride length, speed, etc. This uniqueness in gait patterns is attributed to a combination of anatomical and physiological characteristics that vary from one person to another. Research in biomechanics and human movement science provides the theoretical basis for gait identification, offering insights into how and why gait patterns differ among individuals.

The applications of gait identification are varied, ranging from security and surveillance to health care. In security, gait identification can be used for access control and monitoring in sensitive areas, offering a non-invasive and hard-to-spoof method of identification [2]. In surveillance, it provides a tool for tracking individuals in crowded or public spaces without the need for facial identification or other more intrusive biometrics [3]. Additionally, in the healthcare sector, gait analysis is employed to diagnose and monitor conditions that affect mobility, such as Parkinson's disease or orthopedic injuries, by observing changes in gait patterns over time [4]. Gait can also be analyzed to determine how healthy a person is living. This can be done using a digital assistant [5].

The study of human gait is of paramount importance as it serves as an important indicator of both cognitive and motor stability. Gait analysis is essential to understand how well a person can cope with stress and its potential effects on overall quality of life. In a mobile health (mHealth) sensor model [6], monitoring a person's gait using a smartphone allows for detailed and continuous assessment of their stability. This approach offers a direct assessment of an individual's physical and cognitive health, thereby increasing the ability to support the well-being of frail and elderly people.

One of the factors influencing gait is geographic. In the northern territories, where seasonal changes significantly impact environmental conditions, it is crucial to develop a gait analysis model that is all-season and not limited to summer performance. The distinctiveness of each season—ranging from deep snow and ice in winter to muddy or uneven terrains in spring and fall—poses unique challenges for gait recognition technologies. These conditions can alter a person's walking pattern, making it imperative for the model to adapt and accurately identify individuals across varied terrains and weather conditions. An ability to function reliably throughout the year, especially in extreme northern climates, is essential for the practical application of gait recognition technology in security, medical diagnostics, and assistive technologies, ensuring consistent and accurate performance regardless of seasonal changes. For example, it could be a digital assistant with the help of which a person's motor activity is monitored.

Even though there is a lot of research on this problem, no one has asked the question: "Is it possible to identify a person in winter using a model that was trained on summer data?" This study shows that it is difficult to identify a person in winter using a model trained on a dataset that contains data collected in summer. The main contribution of this study is to create an all-season gait classifier model that can identify a person in both winter and summer.

The rest of the paper is organized as follows. Section

II analyzes related work on similar topics and shows the insufficiency of the existing solution for human identification in winter. Section III describes possible solutions to this problem using smart human sensorics. Section IV describes the collected dataset and the machine learning algorithms used. Section V presents the experimental results. Section VI discusses the usefulness of the findings as well as directions for future work. In conclusion, the results of the study are summed up.

II. Related Work

Person identification by gait involves two main types: video-based identification, which analyzes visual walking patterns, and Inertial Measurement Unit (IMU)-based identification, focusing on sensor-detected motion data. Video identification technique capitalizes on the unique walking patterns of individuals for identification and re-identification, offering a discreet yet effective means of monitoring without requiring direct interaction [7]. Researches include the GaitSet [8] and GaitPart [9] models, which optimize the use of spatial and temporal gait information for improved identification performance.

Related researches in gait analysis predominantly focuses on sensor-based methodologies, indicating a move toward real-world applications outside of laboratory settings. Gait detection accuracy was impressive across a variety of metrics including activity, event and deviation detection, using accelerometers and gyroscopes to improve performance in certain activities [10]. Beyond visual data, researchers have explored radar and Wi-Fi signals for gait identification, employing deep learning to efficiently extract and process distinctive features from these inputs. These advancements underscore the potential of gait analysis in a variety of applications, from security enhancements to anomaly detection, driven by the continuous refinement of algorithms and the integration of diverse data sources [11].

In the same time gait identification faces unique challenges across seasons, particularly when comparing summer and winter conditions. This area, surprisingly understudied, presents significant variations in gait due to environmental factors and clothing differences [12]. In winter, heavy clothing and adverse weather conditions can alter an individual's natural gait, making identification more difficult compared to summer, where lighter clothing and favorable weather conditions allow for more consistent and natural walking patterns. If these factors are not taken into account in the northern territories, then tools configured for gait analysis will not work for almost half of the year. Understanding these seasonal impacts on gait identification is crucial for improving biometric recognition systems. In the researches listed above, there is no study of the influence of the climatic factor on the identification of a person by gait.

In this study, we test the hypothesis that climatic features of northern territories at different times of the year affect the identification of a person by gait. Consequently, models trained to identify people on summer data will not be able to effectively cope with this task in winter. From this we conclude that for effective identification of a person by gait, it is necessary that the training dataset contains data collected in both winter and summer.

III. Gait-Based Personal Identification in Smart Human Sensorics

Smart human sensorics refers to the advanced integration of sensor technology and intelligent systems to monitor, analyze, and interpret human physiological and behavioral data. This field encompasses a wide range of applications, from health monitoring and medical diagnostics to interactive computing and human-machine interfaces. By leveraging sensors that can detect various human parameters such as movement, vital signs, and even emotional states, combined with intelligent algorithms that process and make sense of this data, smart human sensorics aims to enhance our understanding of human health, behavior, and interactions with technology [13].

In the context of health and wellness, smart sensorics can involve wearable devices that track physical activity, monitor heart rate, or assess sleep patterns, providing insights into an individual's health status and alerting them to potential issues before they become serious. In more advanced applications, such as in smart homes or assistive technologies, these sensorics can help adapt the environment to the needs of individuals, improving their quality of life and independence. The integration of artificial intelligence (AI) and machine learning further enhances the capabilities of smart human sensorics, enabling predictive analytics, personalized feedback, and adaptive responses to the unique patterns and needs of individuals [14].

Smart human sensorics and personal identification by gait are interconnected in the realm of advanced biometric systems. Gait analysis, as a form of smart human sensorics, leverages sophisticated sensors and intelligent algorithms to capture and interpret the unique patterns of an individual's walk. This process involves analyzing the complex movements and biomechanical characteristics that are distinct to each person [15]. Such an analysis is especially important for northern territories in winter, when, due to the factors listed above, a person's gait changes significantly. For example, knowing a person's gait patterns in different seasons of the year, we can separate situations where the gait has changed due to a seasonal factor from a situation where the gait has changed due to a progressive illness and the person needs to be sent to the doctor. Applications that implement such functions can be called digital assistants. Also, knowing what time of year the gait data refers to, one can draw conclusions about the sufficiency of a person's motor activity. For example, in winter, when it's cold, people move less actively and this is not associated with poorer health, but depends specifically on the weather and climate.

In the context of personal identification, gait analysis offers a non-invasive, difficult-to-mask method of recognizing individuals based on their walking patterns. This is particularly useful in security and surveillance, where unobtrusive monitoring is required. Integrating smart human sensorics into gait analysis enhances the system's accuracy and adaptability. By utilizing advanced sensors, such as accelerometers, gyroscopes, and pressure sensors, combined with AI-driven data analysis, these systems can accurately capture and analyze gait patterns in real-time. They can adapt to various conditions, such as different walking speeds or changes in terrain, making the identification process more robust and reliable.

Furthermore, as smart sensor technology evolves, it can be integrated into wearable devices, enabling continuous and real-time gait analysis [16]. This can have applications beyond security, including health monitoring, where changes in gait can indicate health issues, and in smart environments, where individuals can be identified and their preferences or needs can be anticipated and addressed automatically. Thus, the fusion of smart human sensorics and gait analysis represents a dynamic and evolving field with a wide array of applications in security, health, and personalized technology interactions.

Incorporating the all-seasons aspect into the context of gait-based personal identification within smart human sensorics emphasizes the necessity for these systems to be versatile and effective regardless of seasonal changes. This is particularly crucial in regions experiencing significant variations in weather and environmental conditions throughout the year, which can influence an individual's gait. For instance, the way a person walks on a snowy winter day differs markedly from their summer stroll, due to factors like footwear, clothing, and the walking surface.

An all-seasons smart human sensorics system must, therefore, have the capability to adapt and maintain high accuracy in personal identification by analyzing gait across diverse conditions. This involves utilizing advanced sensors and algorithms that can discern and adjust to the subtle changes in gait patterns induced by different seasons. Such a system ensures consistent performance whether it's tracking an individual across a slippery, ice-covered pathway in winter or a dry, sandy beach in summer.

Moreover, this all-season functionality enhances the system's applications in various fields, from security, where individuals need to be identified reliably regardless of weather, to healthcare, where monitoring gait changes can provide insights into an individual's physical well-being throughout the year. In essence, integrating all-season adaptability into smart human sensorics for gait identification broadens the scope and utility of this technology, making it a robust tool for personal identification and monitoring in any environmental condition. This study provides a model of human identification that takes into account the seasonality factor.

IV. DATASET AND MACHINE LEARNING ALGORITHMS

A. Hypotheses

1) A person identification model trained on summer data, which identifies a person with high accuracy ($\geq 90\%$) in the summer, will identify a person with low accuracy ($\leq 50\%$) in winter.
2) A model trained on winter data will identify a person by gait on winter data with high accuracy ($\geq 90\%$).
3) A model trained on combined data (the dataset contains both summer and winter data) will identify a person with high accuracy both in summer and winter.

To test these hypotheses, it was necessary to collect data on the gait of people in different seasons of the year. The process of collecting data and testing hypotheses is described in detail below.

B. Sensors

During the experiments, 7 smartphones were placed on subject's body, each of which was equipped with an accelerometer and a gyroscope. The diagram of the location of smartphones on body of a subject is shown in Fig. 1. During the experiment, the following data was collected (all characteristics were collected along three coordinate axes):

- Angular acceleration (ax, ay, az)
- Linear acceleration (gFx, gFy, gFz)
- Angular velocity (wx, wy, wz)

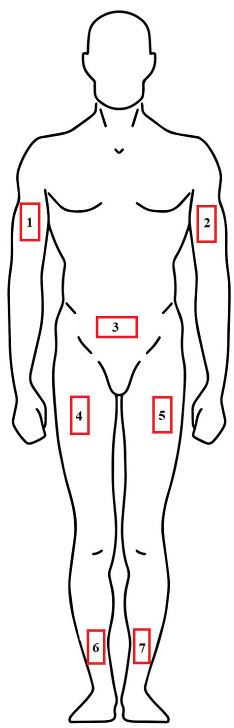

Fig. 1. Scheme of attaching smartphones on body of a subject

The data from the sensors mentioned above were collected using the Physics Toolbox Sensor Suite application, which is available on the App Store and Google Play. This application leverages the smartphone's internal sensors to gather, display, record, and export data files in .csv format. It also provides options to customize the data collection frequency. The brands and models of the phones used for conducting the experiment are listed below. Devices on which the data was recorded comprise 7 phones:

1) Samsung Galaxy S5 - 2 units.
2) Samsung Galaxy S3 Gt19300 - 1 unit.
3) Xiaomi Redmi Note 8 - 1 unit.
4) Xiaomi Note 10 Pro - 1 unit.
5) iPhone 13 - 1 unit.
6) iPhone 14 - 1 unit.

C. Dataset collection

The experiments consisted of two stages (asphalt pavement, dirt road), the total distance included 1200 steps. At each stage of the experiments, the following actions were carried out.

- A subject begins to move at the starting point.
- A subject walks 100 steps in a straight line.
- A subject performs a 180 degree turn.
- A subject walks 100 steps in a straight line.

Factors influencing gait that were taken into account are described in Table I. Weather conditions during data collection are described in Table II. The temperature difference between summer and winter conditions was more than $20°C$. Also, summer data collection was carried out only in natural light, while winter data collection was carried out in both natural and artificial light. In addition, summer data collection took 2 hours less than winter data. This is due to the fact that during winter data collection there was more time spent putting on smartphones. The experimenters also periodically needed to warm up indoor.

TABLE I. FACTORS AFFECTING GAIT

Factor	Description
Season	As part of this study, measurements were taken in September and December 2023
Cloth	The clothing that subjects wore during the experiments was described, but it was not included in the training and testing samples. This is planned to be done in the future
Walking pace	Subjects walked the distance at three paces: fast, normal and slow
Lighting	Subjects walked on the asphalt surface in natural light. Subjects walked on the ground surface in September in natural light, and in December in artificial light
Type of road surface	In September, subjects walked on asphalt and dirt roads. In December, subjects walked through snow on a dirt surface. There was also snow cover on the asphalt surface, but in some places there was icy conditions

Subjects covered this distance 6 times in each season: 3 times on asphalt, 3 times on a dirt road. The need to walk the distance 3 times was caused by the variability of the walking pace. The first time subjects walked quickly, the second time as usual, the third time slowly.

To study the differences in gait taking into account the climatic features of the northern territories, it was necessary to collect a dataset with data on gait in summer and winter. The experiments involved 10 subjects aged from 18 to 28 years, including 7 men and 3 women. Information about each subject is presented in Table III.

TABLE II. WEATHER CONDITIONS

Season	Factor	Value
Summer	Date	September 23
	Start time	9:00
	End time	16:00
	Sunrise	06:28
	Sunset	18:41
	Humidity	64.4%
	Temperature	$17.9°C$
Winter	Date	December 1
	Start time	09:00
	End time	18:00
	Sunrise	09:33
	Sunset	15:30
	Humidity	87%
	Temperature	$-5.1°C$

TABLE III. INFORMATION ABOUT SUBJECTS

№	Sex	Weight (kg)	Height (cm)	Age
1	Female	63	170	20
2	Male	86	183	20
3	Male	83	185	21
4	Female	64	168	20
5	Male	96	193	21
6	Male	89	178	28
7	Male	57	180	21
8	Male	75	170	20
9	Male	80	184	21
10	Female	66	168	18

D. Data filtering

After the dataset was collected, a crucial step in preprocessing involved the application of data filtering to enhance the quality of the data for subsequent analysis. Data filtering is a process used to remove noise and irrelevant information from data, making it more suitable for analysis. In this phase, three primary filters were employed: the median filter, the moving average (MA) filter, and the low-pass Butterworth filter [17]. Each of these filters works on different principles to smooth data and reduce noise. These filters were chosen because they are fast and easy to implement.

The essence of the median filter lies in its ability to replace each data point in the dataset with the median of neighboring points. This non-linear filter is particularly effective in removing "salt and pepper" noise without significantly blurring the edges of the data. It works by sliding a window over the data, calculating the median value within the window, and then replacing the central value with this median. This approach is highly effective in preserving significant data features while eliminating outliers.

The MA filter, a type of linear filter, smooths data by replacing each data point with the average of neighboring data points. A size of the window over which the average is computed can be adjusted based on the desired level of smoothing. This filter is adept at reducing random noise and is often used to reveal underlying trends in the data. However, it tends to blur sharp edges, which may not be desirable in all applications.

The low-pass filter is designed to allow only the low-frequency components of the signal to pass through while attenuating components with frequencies higher than a certain cutoff frequency. This filter is instrumental in removing high-frequency noise and is widely used in signal processing and time series analysis. The essence of the low-pass filter is its ability to preserve the overall shape and trends of the data while eliminating fine-scale fluctuations.

E. Machine learning algorithms

The first method we used was a decision tree (DT). The DT method is a popular machine learning algorithm that models decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is a predictive modeling tool that divides a dataset into smaller subsets while at the same time an associated DT is incrementally developed, enabling both classification and regression tasks by mapping observations about an item to conclusions about the item's target value. In the referenced study [18], the instrumentation comprises shoe-mounted pressure sensors, knee-position encoders, and gyroscopes affixed to the thigh and calf. These devices are employed to quantify the foot's contact force, as well as the angular position and velocity of the knee joint. Subsequently, the C4.5 DT algorithm is utilized to delineate five distinct sub-phases of the walking gait, based on an integrated analysis of the data acquired from these sensors.

The second method we used was the nearest neighbors method. The k-nearest neighbors (k-NN) method is a simple, yet powerful, algorithm used in machine learning for classification and regression tasks, which operates by finding the "k" closest data points in the feature space to a given input point and making predictions based on the majority vote (for classification) or average (for regression) of these neighbors. Its effectiveness is heavily influenced by the choice of "k" and the distance metric used to measure closeness between points. The study [19] explores enhancing behavioral biometrics accuracy, specifically in human gait identification, through ensemble classifiers leveraging k-nearest neighbor algorithms. By analyzing ground reaction forces segmented into gait cycle sub-phases, the method, tested on over 3500 gait cycles from 200 individuals, achieved a correct classification rate exceeding 97.37%.

The third method we used was random forest (RF) method. The RF method is an ensemble learning technique used for classification, regression, and other tasks, which operates by constructing a multitude of DT's at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. This approach offers advantages in terms of accuracy, ability to deal with large data sets with higher dimensionality, and provides measures of feature importance, making it robust against overfitting when compared to single DT. In the study [20] on gait identification for enhancing security in public spaces, the authors employ DT, RF, and k-NN classifiers. Specifically, using public CASIA-A dataset, the RF classifier achieves an 84.26% accuracy, highlighting its effectiveness in pattern identification within the unique context of human gait analysis.

V. RESULTS OF HUMAN IDENTIFICATION BY GAIT

The methodology for data selection employed for model training was executed through three distinct strategies:

1) Summer data. Initially, the model was trained utilizing 80% of the dataset derived from files comprising sensor readings recorded during the summer season. Subsequent testing was conducted on two fronts: first, on the remaining 20% of the summer dataset not utilized in the training phase, and second, on the entirety (100%) of the dataset containing sensor readings acquired in the winter season.
2) Winter Data. In this approach, the model underwent training with 80% of the data extracted from the file encapsulating sensor readings collected during the winter period. The testing phase was conducted on the residual 20% of the winter dataset that was excluded from the training process.
3) Combined data. The model was trained on 80% of the combined datasets, which included sensor readings from both the summer and winter seasons. The testing phase was distinctly organized into three parts: firstly, testing was performed on 20% of the combined datasets not included in the training set. This was followed by separate testing on 20% of the data collected in the summer and not utilized in the training phase, and finally, testing on 20% of the data collected in the winter and excluded from the training set.

These methodologies were designed to evaluate the model's performance under varying training conditions and to assess its generalizability across different seasonal datasets. The performance of decision trees and k-NN algorithms trained on data collected in different seasons was evaluated to understand their effectiveness in generalizing across seasons. Tables IV to XII summarize the results obtained from these experiments, highlighting the impact of seasonal variability and the choice of filtering method on model performance. In the "filter" column, filter parameters are indicated in parentheses. For the median filter, this is the window size equal to the number of dimensions. For a MA, this is the time window within which measurements are averaged. For example, a window of 1 second means that the filter window size is equal to the number of measurements in one second. This is necessary to take into account that the frequency of data collection from different sensors may be different. For the Butterworth filter, the order and sampling frequency (fs parameter) were specified. The operating diagram of the proposed method is shown in Fig. 2.

The results presented in tables IV, VII and X show the results of testing hypothesis №1. Models trained on summer data, when tested on winter data, produced an accuracy of no

TABLE IV. RESULTS OF DECISION TREES TRAINED ON 80% OF THE DATA COLLECTED IN THE SUMMER

Filter	Summer train	Summer test	Winter test
No filter	1.000	0.767	0.138
Median (3)	1.000	0.785	0.136
Median (5)	1.000	0.802	0.136
Median (7)	1.000	0.835	0.138
Median (11)	1.000	0.864	0.136
Median (17)	1.000	0.892	0.137
Median (21)	1.000	0.907	0.134
Median (25)	1.000	0.920	0.134
Median (51)	1.000	0.951	0.127

TABLE V. RESULTS OF DECISION TREES TRAINED ON 80% OF THE DATA COLLECTED DURING THE WINTER PERIOD

Filter	Winter train	Winter test
No filter	1.000	0.800
Median (3)	1.000	0.815
Median (5)	1.000	0.832
Median (7)	1.000	0.865
Median (11)	1.000	0.889
Median (17)	1.000	0.916
Median (21)	1.000	0.932
Median (25)	1.000	0.940
Median (51)	1.000	0.966

TABLE VI. RESULTS OF DECISION TREES TRAINED ON 80% COMBINED DATA

Filter	Summer + winter train	Summer + winter test	Summer test	Winter test
No filter	1.000	0.765	0.752	0.778
Median (3)	1.000	0.785	0.773	0.797
Median (5)	1.000	0.806	0.794	0.818
Median (7)	1.000	0.845	0.832	0.857
Median (11)	1.000	0.874	0.863	0.885
Median (17)	1.000	0.903	0.893	0.913
Median (21)	1.000	0.920	0.910	0.929
Median (25)	1.000	0.931	0.923	0.940
Median (51)	1.000	0.960	0.954	0.966

TABLE VII. RESULTS OF THE K-NEAREST NEIGHBORS METHOD WHEN TRAINED ON 80% OF THE DATA COLLECTED IN THE SUMMER

Filter	Summer train	Summer test	Winter test
No filter	1.000	0.906	0.140
Median (3)	1.000	0.913	0.140
Median (5)	1.000	0.921	0.139
Median (7)	1.000	0.937	0.139
Median (11)	1.000	0.946	0.141
Median (17)	1.000	0.953	0.139
Median (21)	1.000	0.959	0.137
Median (25)	1.000	0.965	0.133
Median (51)	1.000	0.978	0.132

more than 20%. The results presented in Tables V, VIII and XI show the results of testing hypothesis №2. Models trained and tested on winter data, produced an accuracy $\geq 90\%$. The results presented in Tables VI, IX and XII show the results of testing hypothesis №3. Models trained on combined data most often had $\geq 90\%$ accuracy on both summer and winter data. Thus, we can conclude that to create a model for identifying a person by gait, it is necessary to take into account the seasonality factor.

VI. DISCUSSION

For decision trees trained on summer data in table IV a clear trend emerges where the application of median filters improves the model's test accuracy on summer data but has a negligible or adverse effect on its generalization to winter data. Specifically, while the summer test accuracy improves significantly from 0.767 (no filter) to 0.951 (Median(51)), the winter test accuracy decreases from 0.138 to 0.127. This indicates that while filtering improves performance on data similar to the training set, it may not aid in generalizing across significantly different conditions.

A similar trend is observed with decision trees trained on winter data in table V, where winter test accuracy improves from 0.800 to 0.966 with increasing filter size. This reinforces the notion that models tend to perform better on test data that resembles their training data. With random forest trained on winter data in Table XI the test accuracy changes less with increasing filter size.

Combining summer and winter data for training in table VI shows an improvement in generalization, as evidenced by the more balanced performance between summer and winter test datasets. Notably, the application of a Median (51) filter yields the highest test accuracies for both summer and winter datasets, suggesting that a larger dataset encompassing diverse conditions can enhance model robustness.

The k-NN method displays a similar pattern in Table VII to table IX). While the application of median filters con-sistently improves test accuracy on the summer dataset, the improvement is more pronounced in the combined dataset scenario in Table IX. For instance, using a Median (51) filter, the summer and winter test accuracies increase to 0.982 and 0.986, respectively, which are substantial improvements over the no-filter scenario.

Comparatively, the k-NN method outperforms decision trees in the winter test scenario when trained on summer data, suggesting that k-NN may be more resilient to seasonal changes. This could be attributed to the k-NN method's reliance on local data points, which may allow it to adapt better to variations in the data.

The results indicate that both decision trees and k-NN algorithms can benefit from preprocessing techniques such as median filtering to improve performance on the test data. However, the effectiveness of these filters appears to be context dependent, highlighting the need for careful selection based on specific characteristics of training and test data when developing a human gait identification model.

The investigation into the performance of decision trees and k-NN algorithms across different seasonal datasets has underscored a critical aspect of machine learning model development: the necessity of a diverse and comprehensive

TABLE VIII. RESULTS OF THE K-NEAREST NEIGHBORS METHOD TRAINED ON 80% OF THE DATA COLLECTED IN WINTER

Filter	Winter train	Winter test
No filter	1.000	0.900
Median (3)	1.000	0.909
Median (5)	1.000	0.918
Median (7)	1.000	0.937
Median (11)	1.000	0.948
Median (17)	1.000	0.959
Median (21)	1.000	0.966
Median (25)	1.000	0.971
Median (51)	1.000	0.984

TABLE IX. RESULTS OF THE K-NEAREST NEIGHBORS METHOD TRAINED ON 80%COMBINED DATA

Filter	Summer + winter train	Summer + winter test	Summer test	Winter test
No filter	1.000	0.907	0.907	0.898
Median (3)	1.000	0.912	0.915	0.909
Median (5)	1.000	0.922	0.924	0.920
Median (7)	1.000	0.940	0.941	0.939
Median (11)	1.000	0.951	0.950	0.951
Median (17)	1.000	0.960	0.958	0.962
Median (21)	1.000	0.966	0.964	0.969
Median (25)	1.000	0.972	0.970	0.974
Median (51)	1.000	0.984	0.982	0.986

TABLE X. RESULTS OF A RANDOM FOREST TRAINED ON 80% COLLECTED IN SUMMER

Filter	Summer train	Summer test	Winter test
MA (1 sec)	1.000	0.979	0.133
Butterworth (order = 2, fs = 100)	1.000	0.862	0.176
MA (0.5 sec)	1.000	0.973	0.142
Butterworth (order = 3, fs = 100)	1.000	0.858	0.170
MA (0.33 sec)	1.000	0.966	0.153
Butterworth (order = 2, fs = 200)	1.000	0.900	0.169

TABLE XI. RESULTS OF A RANDOM FOREST TRAINED ON 80% COLLECTED IN WINTER

Filter	Winter train	Winter test
MA (1 second)	1.000	0.979
Butterworth (order = 2, fs = 100)	1.000	0.862
MA (0.5 sec)	1.000	0.973
Butterworth (order = 3, fs = 100)	1.000	0.858
MA (0.33 sec)	1.000	0.966
Butterworth (order = 2, fs = 200)	1.000	0.900

training set. The experimental results clearly demonstrate that models trained exclusively on data from a single season exhibit significant limitations in their ability to generalize to data from other seasons. This is particularly evident in the performance disparity between summer and winter datasets.

The application of median filters showed that while preprocessing can improve model accuracy on test data resembling the training set, it does not necessarily aid in overcoming the fundamental challenge of seasonal variability. However, a notable improvement in model generalization was observed when both summer and winter data were included in the training set. This approach enabled the models to achieve higher accuracy across both seasonal test datasets, also highlighting the benefits of a diversified training strategy.

Performance evaluation of a random forest model applied to gait recognition data, emphasizing the impact of different filtering techniques and the diversity of training data on model effectiveness presented in Tables X and XII, showcase the model's accuracy across various scenarios, highlighting the importance of both the choice of filter and the training dataset's composition.

Table X outlines the performance of a random forest trained exclusively on data collected during the summer. The model demonstrates perfect training accuracy across all filtering techniques, indicating a strong fit to the training data. However, the test accuracies reveal significant differences. The MA filter with a 1-second window achieves the highest summer test accuracy (0.979), suggesting that this filter effectively captures relevant features for gait recognition in similar conditions. Yet, the winter test accuracy for this setup is markedly low (0.133), indicating poor generalization to different seasonal conditions.

Butterworth filters, despite their sophisticated design for signal processing, do not consistently outperform the simpler MA filters. For instance, a Butterworth filter with an order of 3 and a sampling frequency of 100 Hz yields lower summer test accuracy (0.862) and marginally improved winter test accuracy (0.176) compared to the MA (1 sec) filter. This pattern suggests that the complexity of the Butterworth filter does not necessarily translate to better performance in this context.

When the random forest is trained on a dataset combining summer and winter data there is a notable improvement in winter test accuracy across all filtering methods. This result is shown in table XII. For example, the MA (1 sec) filter maintains high accuracy in both summer and winter test datasets (0.983 and 0.981, respectively), indicating enhanced generalization capabilities when the model is exposed to diverse training samples.

Interestingly, the Butterworth filter with an order of 2 and a sampling frequency of 200 Hz shows significant improvement in winter test accuracy (0.968) when trained on the combined dataset compared to its performance on the summer-only dataset. This also reinforces the notion that the diversity of the training set plays a crucial role in the model's ability to generalize across different conditions.

The analysis of decision trees and k-NN algorithms trained on seasonal gait data underscores a crucial insight into machine learning model development: the imperative of a diverse

TABLE XII. RESULTS OF A RANDOM FOREST TRAINED ON 80% COMBINED DATA

Filter	Summer + winter train	Summer + winter test	Summer test	Winter test
MA (1 sec)	1.000	0.979	0.983	0.981
Butterworth (order = 2, fs = 100)	1.000	0.848	0.888	0.868
MA (0.5 sec)	1.000	0.973	0.977	0.975
Butterworth (order = 3, fs = 100)	1.000	0.844	0.885	0.864
MA (0.3 sec)	1.000	0.892	0.923	0.908
Butterworth (order = 2, fs = 200)	1.000	0.965	0.970	0.968

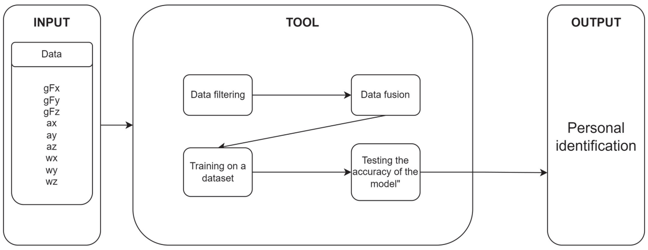

Fig. 2. Scheme of operation of the method of identifying a person by gait

and comprehensive training dataset to enhance model generalizability across different conditions. While median filters improve test accuracy on data similar to the training set, their impact on generalization to distinct seasonal conditions is limited. Models trained solely on summer data struggle to adapt to winter data, indicating a significant challenge in seasonal variability. However, when models are trained on a combined dataset of summer and winter data, there is a noticeable improvement in their ability to generalize, as shown by more balanced accuracies across seasonal test datasets. This finding is pivotal, emphasizing that the inclusion of varied conditions in the training set can significantly bolster a model's robustness and applicability across diverse environmental contexts.

The future of gait analysis is expected to include continuous real-time monitoring and feedback mechanisms to improve rehabilitation and promote quality of life through digital assistant. Despite these advances, the practical implementation of AI in gait analysis as an identification tool remains limited, highlighting the need for further research. The results of this study could become part of a digital assistant that will be used to track a person's physical activity.

VII. CONCLUSION

This study emphasizes the necessity of diverse training datasets for enhancing the generalizability of machine learning models in gait-based identification, demonstrating that models trained on combined seasonal data show improved adaptability to environmental variations, with k-NN, desicion trees and random forest algorithms offering notable resilience to seasonal changes. During the experiments, the following results were obtained:

1) A person identification model trained on summer data, which identifies a person with high accuracy ($\geq 90\%$) in the summer, will identify a person with low accuracy ($\leq 50\%$) in winter.
2) A model trained on winter data identified a person by gait on winter data with high accuracy ($\geq 90\%$).
3) A model trained on combined data (the dataset contains both summer and winter data) will identify a person with high accuracy ($\geq 90\%$) both in summer and winter.

Although the present study was able to achieve an accuracy of 98% in identifying a person in any season of the year, it is not certain that if the sample were increased to a million or more, the accuracy would remain at the same level. Future research should focus on expanding the dataset (primarily due to increasing the number of people in the sample) to cover more nuanced seasonal variations and exploring advanced preprocessing and modeling techniques to further refine the model's performance. Through these efforts, it is possible to achieve a truly adaptable and reliable gait identification system that remains effective regardless of the season.

ACKNOWLEDGMENT

The research is implemented with financial support by Russian Science Foundation, project no. 22-11-20040 (https://rscf.ru/en/project/22-11-20040/) jointly with Republic of Karelia and funding from Venture Investment Fund of Republic of Karelia (VIF RK). The experiment study is performed in collaboration with the Artificial Intelligence Center of PetrSU (https://ai.petrsu.ru/).

REFERENCES

[1] E. J. Harris, I.-H. Khoo, and E. Demircan, "A survey of human gait-based artificial intelligence applications," *Frontiers in Robotics and AI*, vol. 8, p. 749274, 2022.

[2] T. Van hamme, E. A. Rúa, D. Preuveneers, and W. Joosen, "On the security of biometrics and fuzzy commitment cryptosystems: A study on gait authentication," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 5211–5224, 2021.

[3] A. Parashar, A. Parashar, A. F. Abate, R. S. Shekhawat, and I. Rida, "Real-time gait biometrics for surveillance applications: A review," *Image and Vision Computing*, vol. 138, p. 104784, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885623001580

[4] Y. Hutabarat, D. Owaki, and M. Hayashibe, "Recent advances in quantitative gait analysis using wearable sensors: A review," *IEEE Sensors Journal*, vol. 21, no. 23, pp. 26 470–26 487, 2021.

[5] A. Meigal, L. Gerasimova-Meigal, G. Rego, and D. Korzun, "Motor activity sensorics for mhealth support of human resilience in daily life," in *2022 32nd Conference of Open Innovations Association (FRUCT)*, 2022, pp. 169–177.

[6] A. Meigal, L. Gerasimova-Meigal, and D. Korzun, "A concept model of mhealth sensorics for digital assistance of human cognitive resilience," in *2023 34th Conference of Open Innovations Association (FRUCT)*, 2023, pp. 100–107.

[7] E. J. Harris, I.-H. Khoo, and E. Demircan, "A survey of human gait-based artificial intelligence applications," *Frontiers in Robotics and AI*, vol. 8, p. 749274, 2022.

[8] H. Chao, Y. He, J. Zhang, and J. Feng, "Gaitset: Regarding gait as a set for cross-view gait recognition," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 8126–8133.

[9] C. Fan, Y. Peng, C. Cao, X. Liu, S. Hou, J. Chi, Y. Huang, Q. Li, and Z. He, "Gaitpart: Temporal part-based model for gait recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 14 225–14 233.

[10] P. Khera and N. Kumar, "Role of machine learning in gait analysis: a review," *Journal of Medical Engineering & Technology*, vol. 44, no. 8, pp. 441–467, 2020.

[11] V. Rani and M. Kumar, "Human gait recognition: A systematic review," *Multimedia Tools and Applications*, vol. 82, 03 2023.

[12] M. Khaliluzzaman, A. Uddin, K. Deb, and M. J. Hasan, "Person recognition based on deep gait: A survey," *Sensors*, vol. 23, no. 10, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/10/4875

[13] D. Korzun, E. Balandina, A. Kashevnik, S. Balandin, and F. Viola, *Ambient Intelligence Services in IoT Environments: Emerging Research and Opportunities*. IGI Global, 2019. [Online]. Available: https://doi.org/10.4018/978-1-5225-8973-0

[14] D. Bouchabou, S. M. Nguyen, C. Lohr, B. LeDuc, and I. Kanellos, "A survey of human activity recognition in smart homes based on iot sensors algorithms: Taxonomies, challenges, and opportunities with deep learning," *Sensors*, vol. 21, no. 18, p. 6037, 2021.

[15] A. R. Javed, R. Faheem, M. Asim, T. Baker, and M. O. Beg, "A smartphone sensors-based personalized human activity recognition system for sustainable smart cities," *Sustainable Cities and Society*, vol. 71, p. 102970, 2021.

[16] H. Prasanth, M. Caban, U. Keller, G. Courtine, A. Ijspeert, H. Vallery, and J. Von Zitzewitz, "Wearable sensor-based real-time gait detection: A systematic review," *Sensors*, vol. 21, no. 8, p. 2727, 2021.

[17] H. Teh, A. Kempa-Liehr, and K. Wang, "Sensor data quality: a systematic review," *Journal of Big Data*, vol. 7, 02 2020.

[18] Q. Guo and D. Jiang, "Method for walking gait identification in a lower extremity exoskeleton based on c4.5 decision tree algorithm," *International Journal of Advanced Robotic Systems*, vol. 12, no. 4, p. 30, 2015.

[19] M. Derlatka and M. Bogdan, "Ensemble knn classifiers for human gait recognition based on ground reaction forces," in *2015 8th International Conference on Human System Interaction (HSI)*, 2015, pp. 88–93.

[20] A. Madduri, "Human gait recognition using discrete wavelet and discrete cosine and transformation based features," *International Journal of Computer Trends and Technology*, vol. 69, pp. 22–27, 06 2021.