Gender Classification based on Gait Analysis using Ultrawide Band Radar augmented with Artificial Intelligence

Adil Ali Saleem[1], Hafeez Ur Rehman Siddiqui[1*], Rukhshanda Sehar[1], Sandra Dudley[2]

[1] Institute of Computer Science, Khwaja Fareed University of Engineering and Information Technology, Abu Dhabi Road, Rahim Yar Khan, 64200, Punjab, Pakistan; Adilalisaleem@gmail.com; hafeez@kfueit.edu.pk; rukhshandasehar08@gmail.com.

[2] Bioengineering Research Centre, School of Engineering, London South Bank University, 103 Borough Road, London, SE1 0AA, UK; dudleyms@lsbu.ac.uk.

*Correspondence: hafeez@kfueit.edu.pk

Abstract:

The identification of individuals based on their walking patterns, also known as gait recognition, has garnered considerable interest as a biometric trait. The use of gait patterns for gender classification has emerged as a significant research domain with diverse applications across multiple fields. The present investigation centers on the classification of gender based on gait utilizing data from Ultra-wide band (UWB) radar. A total of 181 participants were included in the study, and data was gathered using UWB radar technology. This study investigates various preprocessing techniques, feature extraction methods, and dimensionality reduction approaches to efficiently process UWB radar data. The data quality is improved through the utilization of a two-pulse canceller and discrete wavelet transform. The hybrid feature dataset is generated through the creation of gray-level co-occurrence matrices (GLCM) and subsequent extraction of statistical features. Principal Component Analysis (PCA) is utilized for dimensionality reduction, and prediction probabilities are incorporated as features for classification optimization. The present study employs k-fold cross-validation to train and assess machine learning classifiers, namely Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), and K-Nearest neighbors (KNN). The MLP classifier demonstrates superior performance, with SVM and KNN following closely behind. The findings indicate that the utilization of UWB radar data for gait-based gender classification holds promise in diverse domains, including biometrics, surveillance, and healthcare. The present study makes a valuable contribution to the progress of gender classification systems that rely on gait patterns.

Keywords: Gait analysis, Ultrawide band radar, Gender classification, Principal component analysis, multilayer perceptron, gray-level co-occurrence

1. Introduction:

The term "gait" describes a person's manner of walking. Every person has a distinctive stride that can be used as a behavioral trait to identify them [1]. Gait recognition has become more common in recent years and has a variety of applications, including forensics, access control, surveillance, and the investigation of criminal behavior. Gender classification based on biometric characteristics such as voice, visage, and walking has garnered considerable attention and plays a crucial role in numerous fields [2-6]. It offers numerous benefits, such as robustness in work processes, such as client analysis for administrators and the selection of qualified employees [3]. Unlike traditional recognition methods such as facial, iris, and fingerprint recognition, which rely on physical characteristics or traits, gait recognition utilizes an individual's distinctive walking pattern. This makes it a desirable biometric characteristic, as it overcomes the difficulties of remote identification [7-9]. Therefore, gait recognition has emerged as a significant research field in the field of advanced computer technologies. Thus, identifying human gender based on walking has become an essential area of research.

In recent years, investigators have become increasingly interested in investigating the potential of gender classification based on gait. Significant advancements in this field have been facilitated by the development of sophisticated computer vision algorithms, machine learning (ML) techniques, and the availability of large-scale gait databases. In addition, the pervasive adoption of surveillance systems and the rising demand for non-intrusive identification techniques have fueled the interest in gait recognition for gender classification. This research aims to explore the effectiveness of Ultra-wide band (UWB) radar for gait-based gender classification. This emerging technology provides a non-intrusive and contactless means of capturing gait patterns, enabling accurate gender classification based on unique radar signatures. UWB radar operates by emitting electromagnetic radiation with short pulse durations and measuring the reflected signals [10,11]. UWB radar can capture minute details of human movements, such as gait patterns, by analyzing the time-of-flight and Doppler variations of these signals. This modality offers several advantages over traditional visual-based methods because it is unaffected by lighting conditions, occlusions, or appearance changes [10-13].

The utilization of UWB radar technology for the purpose of gender classification based on gait patterns presents numerous promising prospects. Initially, the non-intrusive characteristic of the UWB radar guarantees [10-14] confidentiality and obviates possible discomfort for the subject's undergoing analysis. Additionally, UWB radar has the capability to function in real-time and under diverse environmental circumstances [15,16], rendering it a viable option for pragmatic implementation in security and surveillance contexts. Finally, the non-intrusive characteristic of UWB radar allows for inconspicuous surveillance of individuals, rendering it suitable for situations where overt collaboration or bodily interaction is unattainable. The findings of this research can have significant implications in security systems, public safety, and other domains where gender-specific analysis is valuable.

The investigation presented in this manuscript offers the subsequent contributions:

- This study investigates the efficacy of UWB radar technology in accurately discerning between male and female gait patterns by gathering radar data from a diverse cohort of individuals at the Khwaja Fareed University of Engineering and Technology (KFUEIT).
- Several preprocessing techniques were utilized to ensure the quality of data and improve the precision of gender classification. The techniques were implemented with the objective of preprocessing the gait data obtained through the utilization of the UWB radar system, in order to facilitate subsequent analysis.
- This study focuses on the development of ML algorithms that are specifically designed to analyze UWB radar data for the purpose of achieving accurate gender classification. The algorithms are developed with a high degree of robustness to ensure reliable results.
- This research incorporates a thorough evaluation of the machine learning algorithms that were developed, with a focus on performance assessment and comparison. Classification performance is evaluated using a range of metrics, including accuracy, precision, recall, F1-score and k-fold cross validation.

2. Literature Review:

The analysis of gait for the purpose of gender classification has been a subject of significant research activity. Numerous studies have investigated diverse methodologies and techniques to attain precise and dependable outcomes. The present literature review examines a chosen set of pertinent studies that enhance comprehension of gender classification based on gait. The review emphasizes the principal findings and methodologies employed in each study. The research conducted by [17] examines the efficacy of a Convolutional Neural Network (CNN) in conjunction with a linear Support Vector Machine (SVM) classifier for the purpose of gender identification, utilizing the CASIA-B dataset. The findings suggest that the SVM exhibits superior performance compared to the Softmax function in the vggnet-16 model for gender classification, resulting in notable accuracy rate of 87.94% across various models.

Research conducted by [18] presents a technique for gender identification utilizing gait data obtained from sensors embedded in smartphones. The Histogram of Gradient (HG) technique is utilized for feature extraction from accelerometer and gyroscope sensors, and a bootstrap aggregating classifier is applied for gender classification that achieved an accuracy of 94.44%. The study presented in [19] proposes a methodology for gait representation that enables the identification and classification of individuals according to their distinctive walking patterns. The study employs the wavelet 5/3 lifting technique to extract gait characteristics from video silhouettes. Additionally, a hybrid approach that utilizes principal component analysis (PCA) and C4.5 algorithm is employed for gender classification. The approach that has been put forth attains classification accuracies of 97.9% and 97.5% when tested on the CASIA-B and OU-ISIR datasets respectively. The research conducted by [20] introduces a gait recognition mechanism that employs statistical characteristics derived from Speed Up Robust Features (SURF). The employed system utilizes various classifiers and attains acceptable levels of accuracy in identifying individuals under different walking conditions, as demonstrated by its performance on the CASIA-B dataset.

The study described in reference [21] employs cost-effective wearable mobile phone sensors to collect gait data for the purpose of gender classification. The study utilizes various machine learning techniques such as SVM, J48, Naive Bayes, and Random Tree to analyze spatiotemporal accelerometer gait data. The Random Tree and Naïve Bayes achieved and accuracy of 83% and 99% respectively. The proposed research [22] presents an integrated methodology for the classification of gait, which involves the combination of multiple classifiers and various gait features. The research gathered data from a diverse group of participants with varying walking patterns, and the suggested approach demonstrated a notable level of accuracy in distinguishing individuals based on their gait. The study presented in reference [23] showcases the utilization of the Kinect sensor in the process of gait identification. The researchers extracted features from joint coordinates and employed machine learning algorithms, including KNN and LDC, that attain an accuracy of 90% and 87.3% respectively.

The utilization of the accelerometer sensor of a smartphone for gender identification based on gait is demonstrated in reference [24]. Multiple characteristics are extracted and a range of machine learning algorithms, such as Decision Tree (DT), SVM, KNN, and RNN-LSTM, are utilized for the purpose of classification. The RNN-LSTM deep learning algorithm attains the utmost accuracy of 94.11% in gender identification. Finally, [25] proposes a pose-based voting technique for gender classification, considering each frame as a labeled instance. The proposed methodology integrates elliptic Fourier descriptors and depth gait histograms, exhibiting a remarkable degree of accuracy even in the presence of partially obstructed gait cycles. The scholarly publication referenced as [26] presents an innovative methodology for forecasting gender and age through the utilization of gait data acquired from inertial sensors. The ML algorithms, specifically KNN, SVM, and DT, are implemented for the purpose of analyzing the extracted time-domain features in order to attain precise gender classification. The SVM model exhibits a notable level of precision, with an accuracy rate of 84.76%, when employed for gender classification. The article referenced as [27] introduces a new methodology for developing a gait recognition system that is impervious to fluctuations in walking velocity. The methodology employed involves the utilizations of Region of Interest (ROI) that is extracted from Gait Energy Image (GEI) for the purpose of classifying probe samples into gallery samples. The categorization of GEI involves the utilization of mutual information to capture its spatial dynamics. The method under consideration was assessed on two publicly accessible gait databases, namely CASIA C and OU-ISIR Treadmill A, and demonstrated a recognition accuracy of 92.66%.

The research presented in reference [28] introduces a CNN model that utilizes gait features to accurately detect gender. The study employs a 3D gait dataset acquired through reflective markers and applies light CNN features with Maximum Feature Maps (MFM) units to mitigate the influence of noisy labels and attain favorable gender detection rates. The CNN framework proposed in this study attains a gender detection rate of 92.7% across all subjects. The article [29] introduces a technique for gender classification that relies on the analysis of gait, with a particular focus on gait patterns that are not neutral. The research employs the CASIA B dataset and introduces a new feature set named Gait Entropy Energy Image (GEnEI) by merging GEI and Gait Entropy Image (GEnI). The efficacy of the proposed model is assessed through the utilization of classification techniques such as KNN and SVM. The AGEnEI and AVGEnEI feature sets exhibit notable classification accuracies of 97.3% and 96.7% correspondingly, upon implementation with SVM. The study [30] introduces a methodology for determining gender and age via gait analysis utilizing a dataset of human gait patterns collected through inertial sensors. Gait cycles are utilized for feature extraction, and a variety of ML algorithms, including Naive Bayes Classifier, Logistic Regression (LR), Multi-Layer Perceptron (MLP), SVM, and RF are utilized for classification purposes. The LR model attains a peak accuracy of 68% in the task of gender classification. The study [31] employs micro-Doppler (m-D) signatures derived from a radar sensor that utilizes frequency modulation in order to identify individuals through analysis of their distinctive gait patterns. The research findings indicate that the identification accuracy of 96.7% was achieved on a test set of 20 subjects by utilizing the distinctive m-D signatures obtained from the radar recordings.

The research conducted in [32] explores the utilization of cost-effective depth cameras in the acquisition of gait features. In this study, the gait data of 81 young participants was collected through the utilization of a dual Kinect V2 depth camera system. The data was then subjected to analysis of various spatiotemporal features in order to facilitate gender classification. The research attains a notable precision rate of 96.7% in gender categorization through the utilization of the identified characteristics. The study conducted by the authors in [33] investigates the potential of utilizing an inertial gait dataset for the purpose of gender classification. The methodology employed in this research involved the application of statistical moments and frequency domain features. Classification is a task that involves the use of machine learning algorithms such as SVM, KNN, bagging, and boosting. The bagging technique attains the utmost precision of 87.858% in the classification of gender. The utilization of gait analysis as a biometric identifier for gender detection is documented in reference [34]. In the study, a dataset of 16 subjects was analyzed to extract 321 gait characteristics. Subsequently, an artificial neural network (ANN) model was trained to detect gender, resulting in a high-level accuracy of 95.83%. The study [35] introduces a framework for the classification of gender and age through the utilization of gait analysis. The study employs a CNN architecture with a tree structure to effectively capture spatiotemporal gait features, resulting in notable accuracies of 97.42% and 99.11% on single-view and multi-view gait datasets, respectively. The study [36] presents a technique for identifying gender through the analysis of gait phases and muscle activity during regular walking. The Recursive Feature Elimination (RFE) technique is employed to identify the optimal features, while the SVM and RF classifiers exhibit notable accuracies of 99.11% and 98.89%, correspondingly, in the task of gender classification.

The authors of [37] have introduced a theoretical framework, known as attention-aware spatial-temporal learning (ASTL), which is utilized for the purpose of gait representation and gender classification. The framework encompasses diverse modules that cater to temporal aggregation, attention aggregation, and multimodal cooperative learning. The method attains a notable level of accuracy in gender classification, with a correct classification rate (CCR) of 97%. The research [38] employs 3D coordinates of diverse joints acquired from a Kinect sensor to classify gender. Significant joints are selected through the use of statistical techniques, followed by the training of a gender classification LR model. The approach attains a notable level of precision in classification, reaching 98.0% accuracy, particularly when utilizing all body joints. The article [39] presents a methodology for gender classification that utilizes lower-body joint data acquired from the Kinect sensor. The utilization of statistical techniques is employed to identify significant correlations, while gender classification is accomplished through the application of a binary logistic regression model. The approach attains a precision rate of 98.3% during real-time experimentation. The study described in reference [40] outlines a methodology for gender classification utilizing real-time, multi-view 3D gait analysis with Microsoft Kinect. The gait data collected from the sensor is utilized to construct a statistical model, which attains a noteworthy accuracy rate of 97.50%.

The aforementioned studies make a collective contribution to the pre-existing knowledge on gender classification based on gait. They investigate various methodologies, datasets, and features to achieve precise gender identification. Numerous obstacles limit the efficacy and generalizability of investigations using gait analysis. The use of datasets that are compact and have an unbalanced distribution raises serious questions. The datasets typically show deficiencies in their ability to capture the wide range of gait patterns that are common in the general population. As a result, the models created from these datasets might only be somewhat generalizable. The fact that gait data is frequently collected in controlled indoor environments raises another problem. Although the controlled setting makes it easier to take exact measurements, it could not accurately reflect real-life scenarios. Uneven terrain, a variety of walking surfaces, and external stimuli are just a few of the factors that can significantly affect someone's gait patterns. However, these aspects are usually not appropriately taken into account throughout the data gathering process, which limits the applicability of the findings to real-world circumstances. In addition, several of the approaches rely on certain pieces of hardware, such as frequency-modulated continuous-wave radar sensors, IMUZ sensors, or depth cameras. Although these tools provide accurate measurements, their use may be limited in real-world situations where their accessibility or practicality are limited. Their limited scope limits the developed approaches' potential for wider use.

3. Methodology:

This technique comprises four fundamental stages. During the initial phase, raw radar data is collected. Subsequently, the subsequent stage involves the removal of unnecessary elements and the feature extraction. Step 3 involves the utilization of ML models for the purpose of classification. ML models were evaluated in step 4. The methodology diagram of the proposed system is shown in Figure 1.

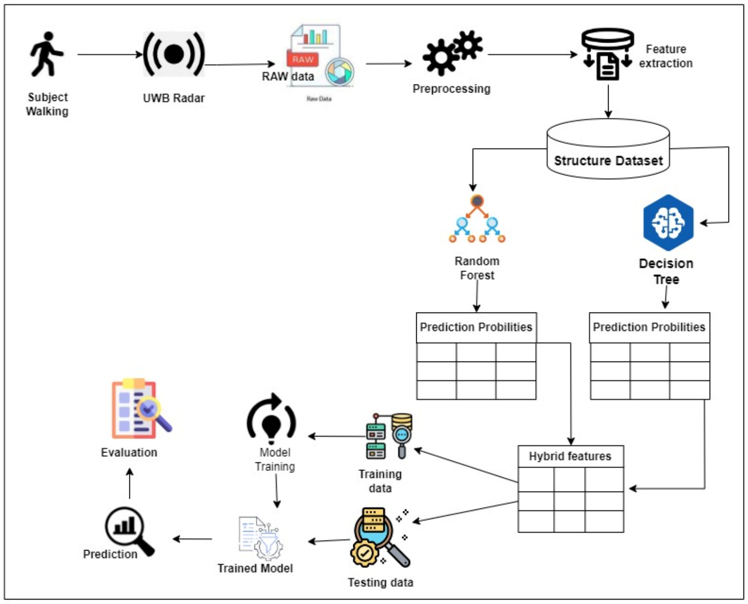

Figure 1: Proposed methodology of the presented system

3.1 Data Collection:

Data was collected from both male and female students at KFUEIT for the study. A total of 181 people (92 males and 89 females) between the ages of 18 and 24 took part in the study. There were a total of 1810 data points gathered from the 10 observations obtained from each participant. The KFUEIT ethics committee gave its stamp of approval after carefully considering the study's ethical implications in light of the Helsinki Declaration. Each participant signed a permission form indicating that they understood the risks and benefits of taking part in the trial. The data was collected through the utilization of a PulseON time domain 410 (P410) Ultra-Wideband (UWB) radar system as shown in Figure 2 encircled in red. The monostatic configuration of the radar system involved the utilization of distinct transmit and receive antennas that were placed in close proximity to each other. The radar range and transceiver were compliant with the guidelines established by the Federal Communications Commission (FCC), emitting radio waves within the 3.1 GHz to 5.3 GHz frequency range, and featuring a central frequency of 4.3 GHz. The operational bandwidth of the system was 2.2 GHz.

The radar system was positioned at a height of 72 cm on a stand, which offered an ideal location for the acquisition of data as shown in Figure 2. The Raspberry Pi (RPi) device was linked to a Virtual Network Computing (VNC) viewer, enabling remote control of radar. The utilization of the VNC viewer facilitated the ability of researchers to remotely access the RPi computer. In the course of data collection, every participant proceeded towards the radar from a distance of 9 meters while maintaining a frontal orientation that can be seen in Figure 2. The duration of data recording for each participant was 5 seconds. Subsequently, the collected data was segregated into distinct directories denoted as "Male" or "Female," classifying the recorded observations according to the gender of the subjects.

Figure 2: Subject moving toward UWB 410 radar mounted on stand and encircled red.

3.2 Preprocessing and Feature Extraction:

In order to improve the accuracy and effectiveness of classifying biomedical data, several essential steps need to be undertaken: preprocessing, feature extraction, and dimensionality reduction [41]. In order to reduce the presence of extraneous information within the data, a two-pulse canceller is employed, as outlined in Equation 1.

$$\mathrm{R}_{\text{output}} = R_i - R_{i-1} \tag{1}$$

Through the use of this equation, the extraneous elements are efficiently eliminated, leading to a modified dataset. The modified matrix identified as 'A'. In order to perform additional operations on matrix 'A', a transformation is applied, resulting in the creation of a grayscale image represented as "I." The process of transformation facilitates the creation of a visual representation of the modified matrix, thereby facilitating the comprehension and evaluation of the information. A single level 2D discrete wavelet transform has been employed to augment the analysis of the grayscale image "I." The conversion results in the production of four distinct sets of coefficients, namely the approximation (LL), horizontal (LH), vertical (HL), and diagonal (HH) detail coefficients. The coefficients offer insights into diverse facets of the image, encompassing multiple frequency components and directional characteristics. In this manuscript only LL and LH coefficients are considered. As the image is grayscale, a grey-level co-occurrence matrix (GLCM) is generated for every decomposed matrix. The GLCM matrices are generated with dimensions of 256 rows and 256 columns and are created at three distinct angles, namely 0 degrees, 45 degrees, and 135 degrees.

The generation of the GLCM matrix at an angle of 0 follows a particular methodology. The matrix element C (i, j) is defined such that 'i' corresponds to the row index and 'j' corresponds to the column index. Specifically, C (0, 0) denotes the frequency of a pixel with a value of 0 being followed by a neighboring pixel with a value of 0 in the rightward direction. The symbol C(0, 1) denotes the occurrence rate of a pixel that possesses a value of 0, which is succeeded by an adjacent pixel that has a value of 1 in the rightward direction. The procedure is iteratively applied to all constituent elements of the GLCM matrix. Upon the completion of the construction of the GLCM matrix, a normalization process is carried out to confine the values within a predetermined range, which is commonly set between 0 and 1. Additionally, statistical features are derived from the normalized GLCM. The characteristics encompass various metrics, namely energy, contrast, correlation, entropy, and homogeneity. The statistical characteristics of an image offer valuable insights into its texture, spatial relationships, and patterns. Combining the extracted features from each decomposed matrix results in a set of twenty-four features. The features, in conjunction with their corresponding labels, are recorded in a comma-separated values (CSV) document. This facilitates the facile preservation, retrieval, and subsequent scrutiny of the characteristics and corresponding image annotations.

Principal component analysis is applied to reduce to the most contributing three features. A 3D scatter plot was plotted using the extracted 3 features as shown in Figure 3.

The analysis of the scatter plot indicated that the extracted features' data points were intricately intertwined, exhibiting a convoluted configuration that posed difficulties in their classification through a ML algorithm. A 3-dimensional scatter plot is a useful visualization tool for examining the interrelationships among three variables, thereby facilitating a deeper understanding of the distribution of the data. The scatter plot data exhibits a dispersed and intersecting pattern, indicating the possibility of an absence of well-defined boundaries or unambiguous distinctions between classes or categories. The circumstance poses a challenge for an ML algorithm to precisely categorize the data solely relying on these characteristics. In order to tackle the issue of heterogeneous and intricate data points, an alternative methodology was implemented with the aim of enhancing the accuracy of prediction. The dataset underwent dimensionality reduction through the application of PCA. The use of this methodology facilitates the identification and extraction of significant patterns and fluctuations within the dataset while simultaneously mitigating the impact of extraneous or duplicative information. According to empirical evidence, incorporating prediction probabilities as supplementary features during model training can lead to better outcomes in terms of prediction accuracy as compared to relying solely on raw data [42]. Consequently, the dataset was utilized to train DT and RF models with max_depth=10 and n_estimators=100 respectively, and subsequently the resultant

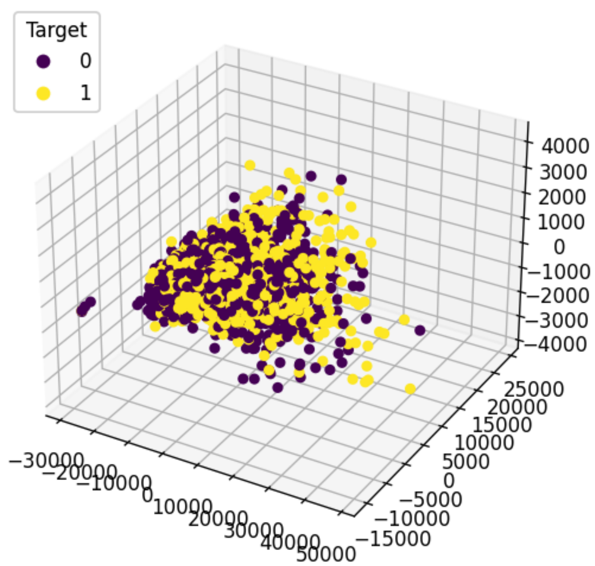

Figure 3: 3D scatterplot of the dataset after dimensionality reduction.

probabilities of prediction were recorded as features. The hybrid feature dataset was merged with its corresponding labels to facilitate further analysis. The hybrid feature dataset, which was generated using prediction probabilities, is visualized through a 3D scatter plot, as illustrated in Figure 4. The scatter plot is noteworthy in that it illustrates the linear separability of the new dataset, thereby suggesting that the data points are more clearly distinguishable and suitable for classification.

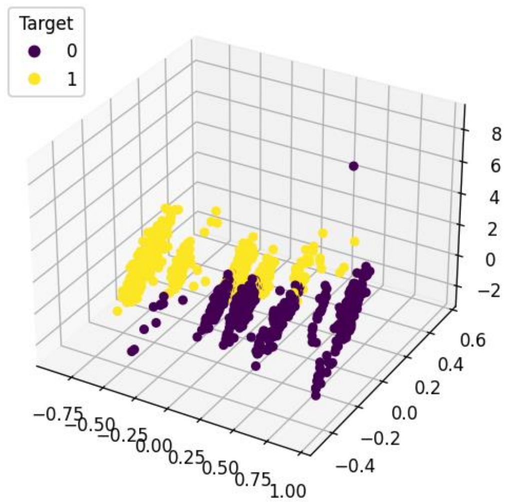

Figure 4: 3D scatter plot of the probability feature dataset.

3.3 *Dataset:*

Each row in the dataset consists of the probabilities assigned by both DT and RF algorithms as features, and the corresponding true gender label (either male or female) as the target or label for gender classification. The dataset comprises of two types of features probability of DT and RF. These features represent the probability or confidence score assigned by a DT and RF algorithm for classifying the gender of a given sample. The snippet of the dataset is given in Figure 5.

0	0.612069	0.387931	0.562100	0.437900	Females
1	0.772727	0.227273	0.589312	0.410688	Females
2	0.851064	0.148936	0.521712	0.478288	Females
3	0.230769	0.769231	0.630103	0.369897	Females
4	0.612069	0.387931	0.525016	0.474984	Females
...	...	...	...	...	...
1805	0.416667	0.583333	0.423541	0.576459	Males
1806	0.439394	0.560606	0.452817	0.547183	Males
1807	0.000000	1.000000	0.482378	0.517622	Males
1808	0.510204	0.489796	0.470234	0.529766	Males
1809	0.531250	0.468750	0.384029	0.615971	Males

1810 rows × 5 columns

Figure 5: Snippets of probability dataset.

The cubic scatter plot is given in Figure 6. To visualize high-dimensional data in a lower-dimensional domain, cubic scatter plots may be made using the Hyper Tools Python module. It accomplishes this by using the dimensionality reduction approach known as t-SNE (t-Distributed Stochastic Neighbor Embedding). Building a probability distribution that captures the similarity between pairs of high-dimensional data points is the fundamental idea behind t-SNE. The t-SNE algorithm uses a Gaussian distribution centered on each data point to construct this probability distribution. The similarity between data points in the lower-dimensional space is captured by this distribution. Finding a mapping from the high-dimensional space to the lower-dimensional space while maintaining the data's natural structure is the goal of t-SNE. The goal of t-SNE is to provide a visualization in which comparable data points are clustered together closely and dissimilar data points are separated by greater distances by modifying the placements of data points in the lower-dimensional space. It is evident from the cubic scatter plot given in Figure 6 that the data is linearly separable. The dataset was used to train and test multiple machine learning classifiers. The findings of these experiments are outlined in the results section, offering valuable perspectives on the efficacy and proficiency of various classifiers in precisely forecasting the gender relying on the hybrid feature dataset.

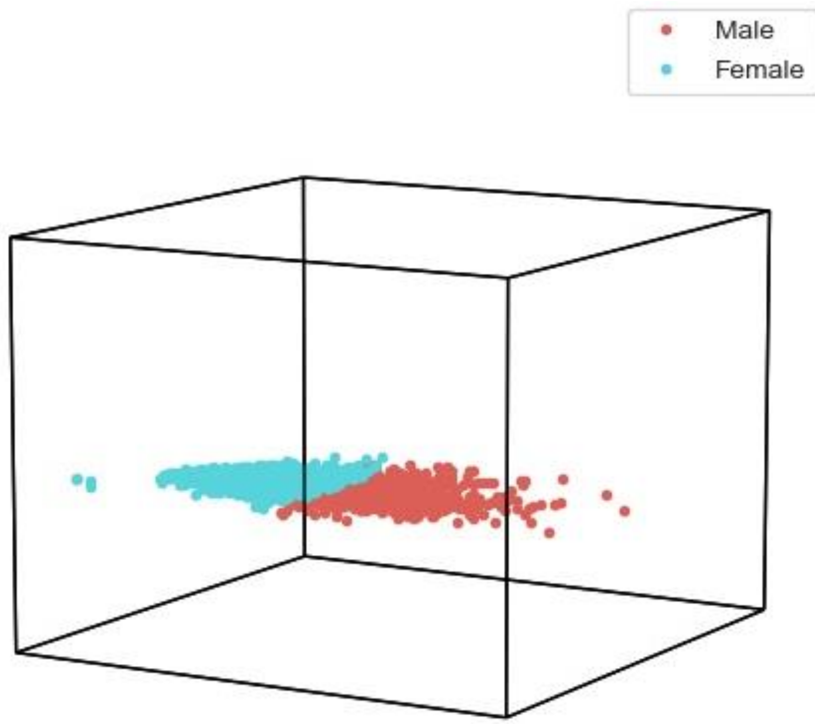

Figure 6: Cubic Scatter plot of the dataset.

4. Results:

The probability dataset from the earlier phases was divided into training and testing sets in the study with a 70:30 ratio. The partitioning scheme implemented ensures that 70% of the available data is allocated for the purpose of training the machine learning models, while the remaining 30% is set aside for the purpose of assessing their efficacy. The experiment employed various machine learning classifiers, namely Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Multi-Layer Perceptron (MLP), K-Nearest Neighbors (KNN), and Extra Tree Classifier (ETC). The previously mentioned classifiers embody a heterogeneous array of computational techniques, each possessing unique attributes and capabilities. The performance of the classifiers was optimized through the tuning of hyperparameters using grid search. The grid search method is a systematic approach that explores a predetermined parameter space in order to identify the most favorable hyperparameter configuration for a given model. The methodology entails the establishment of a set of values for every hyperparameter, followed by a comprehensive assessment of all feasible permutations. Table 1 presents the hyperparameters that were determined to be appropriate for each classifier using grid search.

Table 1: Hyperparameters used to tune the ML classifiers.

Classifier	Hyperparameters
DT	max_depth=200
LR	solver='saga',multi_class='multinomial',C=1.0
KNN	n_neighbors=5,leaf_size=35, p=2
RF	n_estimators=500, max_depth=10
SVM	kernel='poly', C=4.0, probability=True
ETC	n_estimators=300, max_depth=300
MLP	hidden_layer_sizes=(100,50), activation='relu', solver='adam'

The ML classifiers were trained on a subset of 70% of the available data, with the remaining 30% reserved for the purpose of testing and evaluating their performance. This partitioning scheme facilitates an evaluation of the extent to which the models exhibit generalizability to novel data and affords valuable perspectives on their predictive proficiencies. Table 2 displays and Figure 7 visualizes the classification matrix for the various models, offering a comprehensive analysis of the forecasts generated by each classifier.

Table 2: Classification matrices of ML classifiers.

Classifier	Accuracy	Precision	Recall	F1_Score
DT	0.915	0.92	0.92	0.92
LR	0.908	0.91	0.91	0.91
KNN	0.934	0.93	0.93	0.93
RF	0.928	0.93	0.93	0.93
SVM	0.934	0.93	0.93	0.93
ETC	0.91	0.91	0.91	0.91
MLP	0.936	0.94	0.94	0.94

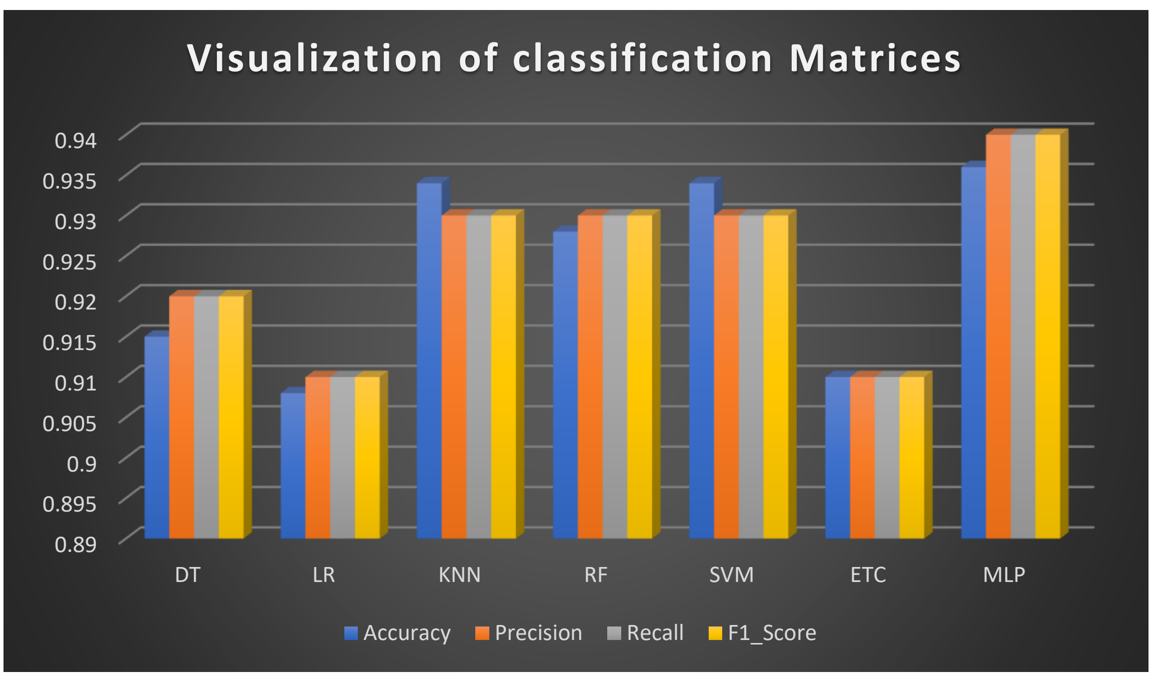

Figure 7: Visualization of classification matrices.

It is evident from Table 2 and Figure 7 that the classifiers exhibited a relatively consistent level of performance across the various metrics. MLP achieves the highest accuracy (0.936), precision (0.94), recall (0.94), and F1 score (0.94) among all classifiers, indicating its overall strong performance. KNN, RF, and SVM also demonstrate high accuracy, precision, recall, and F1 scores, with values of 0.934 across all metrics, reflecting their consistent and balanced performance. DT achieves an accuracy of 0.915, precision of 0.92, recall of 0.92, and an F1 score of 0.92, while LR and ETC achieve slightly lower but still satisfactory results in all metrics, with values of 0.908 and 0.91, respectively.

K-fold cross-validation is a technique commonly used in machine learning to assess the performance of a model in a more robust and reliable manner. It involves splitting the dataset into K equal-sized folds, where K is typically chosen as 5 or 10 here in this manuscript K is 20. The use of K-fold cross-validation enables a more resilient evaluation of classifiers through the iterative assessment of their performance on distinct subsets of the data. The tabulated findings in Table 3 offer valuable insights into the average efficiency and consistency of the models.

Table 3: Cross validation results.

Classifier	**Accuracy**
DT	0.90±0.03
LR	0.90±0.03
KNN	0.93±0.03
RF	0.91±0.03
SVM	0.93±0.03
ETC	0.90±0.03
MLP	0.93±0.03

Due to its high average accuracy and low variability, the MLP appears to be the most promising classifier for the gender classification task based on these cross-validation and testing results given in Table 2 and 3. Nevertheless, KNN and SVM classifiers exhibit comparable efficacy and stability.

5. Conclusion:

The utilization of UWB radar data for gait-based gender classification demonstrates a promising method for precise gender prediction based on gait patterns. The application of preprocessing procedures, feature extraction methods, and dimensionality reduction techniques, in conjunction with the implementation of a two-pulse canceller and discrete wavelet transform, facilitates proficient data processing. A hybrid feature dataset is generated through the creation of GLCM matrices and the subsequent extraction of statistical features. By means of principal component analysis (PCA) and incorporating prediction probabilities as features, the dataset has been optimized for classification purposes. Several machine learning classifiers were trained and tested, and the MLP model was found to be the highest performing, with SVM and KNN following closely behind. The use of K-fold cross-validation serves to validate the robustness of the models. The present study exhibits potential implications in the domains of biometrics, surveillance, and healthcare. Further investigation could augment the precision and practicality of the gender classification mechanism that relies on gait patterns utilizing UWB radar data. The forthcoming research in gender classification based on gait using UWB radar data encompasses various domains of concentration. The proposed approach involves the acquisition of extensive and varied datasets, utilization of data augmentation methodologies, and investigation of sophisticated feature engineering methods to enhance the overall generalization and classification efficacy of the system.

Author contributions:

Adil Ali Saleem: Data curation, Visualization, Investigation, Methodology, Writing- Original draft preparation**, Hafeez Ur Rehman Siddiqui:** Conceptualization, Supervision, Methodology, Software**,** Formal analysis, **Rukhshanda Sehar:** Data curation, Visualization**, Sandra Dudley:** Supervision, Writing- Reviewing and Editing, Validation

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of KFUEIT (HREC-22-690, 20th -June-2022).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.
Data Availability Statement: Data will be provided on demand.

Conflicts of Interest: The authors declare no conflict of interest.

References:

1. Upadhyay, J.; Gonsalves, T. Robust and Lightweight System for Gait-Based Gender Classification toward Viewing Angle Variations. *AI* **2022**, *3*, 538-553.
2. Piccardi, M. Background subtraction techniques: a review. 2004; pp. 3099-3104.
3. Yu, S.; Tan, T.; Huang, K.; Jia, K.; Wu, X. A study on gait-based gender classification. *IEEE Transactions on image processing* **2009**, *18*, 1905-1910.
4. Arai, K.; Andrie, R. Gender classification with human gait based on skeleton model. 2013; pp. 113-118.
5. Harb, H.; Chen, L. Gender identification using a general audio classifier. 2003; pp. II-733.
6. Lam, L.; Lee, S.-W.; Suen, C.Y. Thinning methodologies-a comprehensive survey. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **1992**, *14*, 869-885.
7. Nordin, M.D.J.; Saadoon, A. A survey of gait recognition based on skeleton model for human identification. *Research Journal of Applied Sciences, Engineering and Technology* **2016**, *12*, 756-763.
8. Nixon, M.S.; Carter, J.N. Automatic recognition by gait. *Proceedings of the IEEE* **2006**, *94*, 2013-2024.
9. Maray, A.H.; Hasan, S.Q.; Mohammed, N.L. Design and implementation of low-cost vein-viewer detection using near infrared imaging. *Indonesian Journal of Electrical Engineering and Computer Science* **2022**, *29*, 1039-1046.
10. Siddiqui, H.U.R.; Saleem, A.A.; Brown, R.; Bademci, B.; Lee, E.; Rustam, F.; Dudley, S. Non-invasive driver drowsiness detection system. *Sensors* **2021**, *21*, 4833.
11. Siddiqui, H.U.R.; Saleem, A.A.; Bashir, I.; Zafar, K.; Rustam, F.; Diez, I.d.l.T.; Dudley, S.; Ashraf, I. Respiration-Based COPD Detection Using UWB Radar Incorporation with Machine Learning. *Electronics* **2022**, *11*, 2875.
12. Peng, P.; Yu, C.; Xia, Q.; Zheng, Z.; Zhao, K.; Chen, W. An indoor positioning method based on UWB and visual fusion. *Sensors* **2022**, *22*, 1394.
13. Yang, B.; Li, J.; Zhang, H. Resilient indoor localization system based on UWB and visual–inertial sensors for complex environments. *IEEE Transactions on Instrumentation and Measurement* **2021**, *70*, 1-14.
14. Hu, J.; Jiang, H.; Liu, D.; Xiao, Z.; Dustdar, S.; Liu, J.; Min, G. BlinkRadar: Non-Intrusive Driver Eye-Blink Detection with UWB Radar. 2022; pp. 1040-1050.
15. Liang, X.; Deng, J.; Zhang, H.; Gulliver, T.A. Ultra-wideband impulse radar through-wall detection of vital signs. *Scientific reports* **2018**, *8*, 13367.
16. Zhang, C.; Kuhn, M.J.; Merkl, B.C.; Fathy, A.E.; Mahfouz, M.R. Real-time noncoherent UWB positioning radar with millimeter range accuracy: Theory and experiment. *IEEE Transactions on Microwave Theory and Techniques* **2009**, *58*, 9-20.
17. Liu, T.; Ye, X.; Sun, B. Combining convolutional neural network and support vector machine for gait-based gender recognition. 2018; pp. 3477-3481.
18. Jain, A.; Kanhangad, V. Gender classification in smartphones using gait information. *Expert Systems with Applications* **2018**, *93*, 257-266.
19. Hassan, O.M.S.; Abdulazeez, A.M.; TİRyakİ, V.M. Gait-based human gender classification using lifting 5/3 wavelet and principal component analysis. 2018; pp. 173-178.
20. Htun, K.Z.; Zaw, S.M.M. Human identification system based on statistical gait features. 2018; pp. 508-512.
21. Nutakki, C.; Edakkepravan, H.; Gunasekaran, S.; Ramachandran, L.P.; Sasi, V.; Nair, B.; Diwakar, S. Torque analysis of male-female gait and identification using machine learning. 2018; pp. 2103-2106.
22. Sun, L.; Yuan, Y.-X.; Zhang, Q.; Wu, Y.-C. Human gait classification using micro-motion and ensemble learning. 2018; pp. 6971-6974.
23. Ahmed, M.H.; Sabir, A.T.; Maghdid, H.S. Kinect-based human gait recognition using triangular gird feature. 2018; pp. 1-6.
24. Sabir, A.T.; Maghdid, H.S.; Asaad, S.M.; Ahmed, M.H.; Asaad, A.T. Gait-based gender classification using smartphone accelerometer sensor. 2019; pp. 12-20.

25. Isaac, E.R.H.P.; Elias, S.; Rajagopalan, S.; Easwarakumar, K.S. Multiview gait-based gender classification through pose-based voting. *Pattern Recognition Letters* **2019**, *126*, 41-50.
26. Khabir, K.M.; Siraj, M.S.; Ahmed, M.; Ahmed, M.U. Prediction of gender and age from inertial sensor-based gait dataset. 2019; pp. 371-376.
27. Anusha, R.; Jaidhar, C.D. An approach to speed invariant gait analysis for human recognition using mutual information. 2019; pp. 1616-1621.
28. Wazzeh, A.; Birdal, R.G.; Sertbaş, A. Human Gait Based Gender Detection Using Light CNN with Max Feature Map. 2019; pp. 1-4.
29. Mawlood, Z.Q.; Sabir, A.T. Human gait-based gender classification using neutral and non-neutral gait sequences. *Innovaciencia* **2019**, *7*, 1-13.
30. Gillani, S.I.; Azam, M.A.; Ehatisham-Ul-Haq, M. Age estimation and gender classification based on human gait analysis. 2020; pp. 1-6.
31. Ni, Z.; Huang, B. Human identification based on natural gait micro-Doppler signatures using deep transfer learning. *IET Radar, Sonar & Navigation* **2020**, *14*, 1640-1646.
32. Guffanti, D.; Brunete, A.; Hernando, M. Non-invasive multi-camera gait analysis system and its application to gender classification. *IEEE Access* **2020**, *8*, 95734-95746.
33. Pathan, R.K.; Uddin, M.A.; Nahar, N.; Ara, F.; Hossain, M.S.; Andersson, K. Gender classification from inertial sensor-based gait dataset. 2021; pp. 583-596.
34. ŞEntÜRk, Z.K. Gait Data for Efficient Gender Recognition. *Avrupa Bilim ve Teknoloji Dergisi* **2021**, 27-31.
35. Lau, L.K.; Chan, K. Tree structure convolutional neural networks for gait-based gender and age classification. *Multimedia Tools and Applications* **2023**, *82*, 2145-2164.
36. Lee, M.; Lee, J.-H.; Kim, D.-H. Gender recognition using optimal gait feature based on recursive feature elimination in normal walking. *Expert Systems with Applications* **2022**, *189*, 116040.
37. Huang, B.; Luo, Y.; Xie, J.; Pan, J.; Zhou, C. Attention-aware spatio-temporal learning for multi-view gait-based age estimation and gender classification. *IET Computer Vision* **2022**.
38. Azhar, M.; Ullah, S.; Ullah, K.; Syed, I.; Choi, J. A Gait-Based Real-Time Gender Classification System Using Whole Body Joints. *Sensors* **2022**, *22*, 9113.
39. Azhar, M.; Ullah, S.; Ullah, K.; Rahman, K.U.; Khan, A.; Eldin, S.M.; Ghamry, N.A. Real-Time Dynamic and Multi-View Gait-Based Gender Classification Using Lower-Body Joints. *Electronics* **2022**, *12*, 118.
40. Azhar, M.; Ullah, S.; Raees, M.; Rahman, K.U.; Rehman, I.U. A real-time multi view gait-based automatic gender classification system using kinect sensor. *Multimedia Tools and Applications* **2023**, *82*, 11993-12016.
41. Abdulhamit, S. Practical guide for biomedical signals analysis using machine learning techniques. **2019**.
42. Raza, A.; Siddiqui, H.U.R.; Munir, K.; Almutairi, M.; Rustam, F.; Ashraf, I. Ensemble learning-based feature engineering to analyze maternal health during pregnancy and health risk prediction. *Plos one* **2022**, *17*, e0276525.