and

Assessment of the Biomechanical Clinical Parameters Involved in Pressure Ulcers in Diabetic Neuropathic Patients

August 3, 2025

Internship Report

Helya AMIRI

Supervised by:
Pierre-Yves.ROHAN
jerome HADDAD

Contents

Abstract

Management of the diabetic neuropathic foot represents a major public health challenge. Despite improvements in knowledge regarding ulcer prevention, increased preventive techniques, and the use of new technologies, foot ulcers remain too frequent. Tissue deformation has proven to be an effective mechanical biomarker to predict the risk of developing plantar ulcers.

The objective of this pilot study involving 21 neuropathic diabetic patients, is to propose a simple protocol to estimate internal tissue stresses based on individual parameters accessible in routine clinical practice, to explore correlations among these parameters, and to identify patient groups in order to deduce a new ulcer risk grading system. The proposed protocol is designed to evaluate bone geometry, tissue morphology, mechanical properties of tissues, and tissue perfusion.

This report presents a comprehensive data-driven analysis of the DIAFOOT dataset, which integrates biomechanical, clinical, vascular, and thermographic parameters collected from neuropathic diabetic patients. An interactive analytics pipeline was developed using Python and deployed as a Streamlit dashboard, enabling dynamic exploration, visualization, and statistical evaluation of the dataset.

The results highlight tissue-level markers and biomechanical asymmetries with strong discriminatory power between risk groups, supporting the development of targeted prevention strategies. This work demonstrates the potential of integrated biomechanical-clinical analytics to enhance diabetic foot risk assessment.

1 Introduction

1.1 Scientific and Medical Background

Diabetes is a chronic disease characterized by an excess of sugar in the blood, known as hyperglycemia; prolonged hyperglycemia exposes patients to various complications. Type 2 diabetes, which is the most common form, accounting for over 90% of cases and is rapidly increasing globally due to poor diet and physical inactivity. In France, the number of adults living with diabetes increased from 3.2 million in 2011 to approximately 4.1 million in 2024.(World Health Organization, 2016; International Diabetes Federation, 2024). Approximately 10% of these patients are at risk of lower limb amputation, with more than 10,000 amputations performed each year. These amputations typically stem from infections in chronic wounds, defined as wounds that fail to heal within 4–6 weeks. Diabetes induces two major complications: peripheral neuropathy, which leads to loss of protective sensation, hyperkeratosis, subcutaneous bleeding, and ultimately skin ulceration. The so-called “mal perforant” most commonly appears on the plantar surface of the foot, though similar mechanisms occur in areas exposed to pressure or friction. Treating these chronic wounds poses a major public health challenge.

In this context, the skin’s barrier function is severely compromised. In response, overlapping sequential processes—hemostasis/coagulation, inflammation, proliferation, and remodeling—are activated to restore tissue integrity. Keratinocytes are critical for re-epithelialization, rebuilding the epidermis and barrier function. Reactive oxygen species (ROS) are also key mediators, promoting cell proliferation, migration, and angiogenesis.

Mechanical trauma is considered a key factor in the development of many diabetic foot ulcers (Cavanagh and Bus, 2010). A high plantar pressure peak, often measured to quantify focal zones of mechanical trauma, is a major risk factor for ulceration. Conversely, it has been shown that redistributing plantar pressure away from these high-risk areas—commonly referred to as "offloading"—leads to faster healing of uncomplicated neuropathic foot ulcers (Paton et al., 2011; Raspovic and Landorf, 2014). Consequently, offloading is recommended as the primary therapy for both prevention and management of such ulcers. Offloading interventions, such as the use of custom-made orthopedic insoles or total contact casts, are recommended for reducing pressure and preventing ulcer recurrence. However, adherence to offloading practices can be challenging. Despite improved knowledge in wound prevention, an increase in preventive strategies, and the use of new technologies, foot ulcers remain too frequent (Boulton et al., 2005).

Numerous guidelines have emphasized the necessity of conducting an initial assessment to identify patients at risk of developing plantar ulcers (O’Tuathail and Taqi, 2011). A systematic review published in 2013 identified reduced mobility as the primary risk factor, associated with various patient-specific characteristics including skin condition, perfusion, hydration, and nutritional status (Coleman et al., 2013). Despite this knowledge, both the clinical and scientific communities continue to face challenges in establishing individualized prevention strategies due to the wide diversity and complex interactions of risk factors (Coleman et al., 2013).

From a biomechanical perspective, several studies have shown that: (1) maximum internal shear strain, particularly as defined by the Green–Lagrange strain formulation, and (2) the duration of external loading, are effective biomechanical biomarkers for predicting ulcer risk, at least in animal models using indentation testing (Ceelen et al., 2008; Loerakker et al., 2010, 2011; Gefen et al., 2021; Flynn, 2010). However, in humans, in vivo estimation of such internal tissue stresses remains a major challenge. Computational methods, particularly finite element modeling, offer promising approaches to estimate these internal stresses. The IBHGC institute has been actively developing such methods over the past several years (Keenan et al., 2022).

The subject-specific anatomy of individuals is clearly identified in the literature as one of the key factors

influencing the relationship between external pressures and internal deformations. However, this critical anatomical variability is not adequately considered in current preventive methods. Yet it is precisely these internal deformations that are most closely linked to the ulceration risk faced by patients.

There is therefore a clear clinical need for a simple method to estimate these internal mechanical stresses, based on individual parameters that are accessible in routine clinical settings.

1.2 Problem Statement

Diabetic foot ulcers (DFUs) are among the most severe complications of diabetes, associated with high morbidity, increased risk of lower-limb amputation, and substantial healthcare costs Armstrong et al. (2017); Hingorani et al. (2016); Jeffcoate and Harding (2003). Critically, current clinical tools struggle to identify patients at highest risk before ulceration occurs, as they typically rely on limited measures—such as vascular assessment, neuropathy screening, and surface plantar pressure—and do not fully capture the complex interplay of internal biomechanics, tissue properties, and systemic physiology Bus et al. (2016); Mens et al. (2023); Naemi et al. (2017).

Recent evidence suggests that ultrasound-based measurements of plantar soft tissue stiffness and thickness significantly improve DFU risk prediction: one study showed adding these variables increased sensitivity by 14%, specificity by 3%, and overall accuracy by 5% compared to models using clinical parameters alone Li et al. (2025). Moreover, structural changes like increased tissue stiffness and thickness have been observed in diabetic patients—particularly those with neuropathy or ulceration—highlighting their relevance to ulcer development Naemi et al. (2016); Bus et al. (2023).

Despite IWGDF risk stratification systems being valuable tools, their grading lacks granularity and validation; for instance, up to six risk systems show similar core variables but diverge in predictive accuracy, and the IWGDF system has limited ability to account for biomechanical variability. Unsupervised learning techniques, like clustering, offer a promising avenue to reveal underlying patterns and patient subgroups based on quantitative biomechanical and clinical data, potentially uncovering new, clinically meaningful stratification models.

1.3 Objectives

This study aims to improve the prediction and prevention of diabetic foot ulceration by integrating biomechanical and clinical parameters into more precise risk models. Specifically, the objectives are to address the following key questions and methodologies:

- To identify and validate key predictive markers (e.g., tissue stiffness, thickness, plantar pressure, $TcPO_2$, TBI) that strongly correlate with ulceration risk, contributing to simplified and robust screening protocols.
- To explore the relationship between structural tissue properties and dynamic plantar loading during gait, and assess left–right asymmetries as localized indicators of risk.
- To evaluate the potential of unsupervised learning (e.g., clustering) to stratify patients based on biomechanical and clinical profiles.
- To compare data-driven clusters with existing IWGDF grades and assess whether clustering reveals more physiologically relevant or personalized subgroupings.
- To address the following key questions:

1. Can we identify distinct patient clusters based solely on comprehensive biomechanical and clinical foot parameters—without relying on existing IWGDF categories?
2. Do these clusters align with IWGDF risk grades, or do they reveal novel subgroup structures?
3. Which parameters distinguish transitions between risk levels (e.g., IWGDF 0 → 1 or 2, 2 → 3), and can these insights inform a more precise, clinically actionable risk model?

Hypotheses to Be Tested

- **Hypothesis 1 (Left–Right Asymmetry) – Null Hypothesis:** There is no significant difference between the left and right feet in biomechanical or clinical measures for patients with diabetic neuropathy. In other words, any observed left–right differences (e.g., in plantar pressure, joint range, tissue stiffness, blood flow) occur by chance.
- **Hypothesis 2 (Clustering vs. IWGDF Stratification) – Null Hypothesis:** Any clusters identified in the combined clinical and biomechanical dataset are consistent with, or explainable by, existing IWGDF risk grades. That is, clustering adds no new stratification beyond what is already captured by IWGDF categories.

Note: The IWGDF's emphasis on structured risk grading based on clinical criteria (e.g., mild, moderate, or severe neuropathy) is documented in the literature and serves as the reference stratification model against which new data-driven groupings can be compared.

2 Materials and Methods

This study was conducted in a day-hospital setting in accordance with ethical standards and was registered under the MR-004 declaration. All participants received detailed information about the study and provided written informed consent prior to inclusion. Eligible participants were adults aged 18 to 65 years, of either sex, with a confirmed diagnosis of diabetes, and enrolled in the French national health insurance system. Individuals were excluded if they were pregnant, breastfeeding, or under legal protection measures such as guardianship or trusteeship.

2.1 Dataset Description (DIAFOOT)

Data were collected from 21 participants with valid gender information, including 11 females (52.4%) and 10 males (47.6%). The average age was 56.8 years (SD = 14.1), and the mean Body Mass Index (BMI) was 25.83 ± 5.29 kg/m^2 (range: 18.13–35.00). The BMI distribution showed quartiles at 21.65 (Q1), 25.49 (median), and 30.30 (Q3).

Participants included both diabetic patients. Among diabetic individuals, the cohort spanned a spectrum of disease durations and types: Type 1 (n = 7, mean age of onset = 28.3 ± 8.8 years), Type 2 (n = 11, mean age of onset = 15.4 ± 8.5 years), and Atypical diabetes (n = 1, age of onset = 12 years).

IWGDF (International Working Group on the Diabetic Foot) risk grades ranged from 0 to 3, allowing assessment across the full spectrum of ulceration risk:

- **Grade 0 (Low risk)**: No loss of protective sensation (LOPS) and no peripheral artery disease (PAD),
- **Grade 1 (Moderate risk)**: LOPS or PAD, without deformity or history of ulcer,
- **Grade 2 (High Risk)**: LOPS or PAD, with associated deformity and/or AOMI,

- **Grade 3 (Very high risk)**: LOPS or PAD with active ulcer, amputation history, or both.

Demographic, clinical, and biological data, along with results from quantitative assessments (e.g., plantar pressure, ultrasonography, durometry), were systematically recorded using an electronic Case Report Form (eCRF).

2.2 Participants and Protocol

First, the B-mode US images were collected in 3 locations of the plantar foot region: hallux, medial sesamoid, and 5th metatarsal head and for geometry and material properties. Second, a commercially available pressure sensor was used to collect pressure parameters at a self-selected speed during barefoot walking for 5 trials. The alternative Toe Brachial Pressure Index Systems for a Peripheral Arterial Disease screening was used to check the agreement. Assessment of stiffness, geometry, and tissue perfusion parameters using routine clinical measurement tools in the plantar region of the foot. The demographic, clinical, and biological data of the included subjects as well as the quantitative results were reported on the electronic case report form (e-CRF).

Summary of Ultrasound, Pressure, and Microvascular Protocols

A concise overview of the data collection protocols is provided below, with parameter estimation summarized. All procedures were conducted or CHU Pitié Salpêtrière, Paris, France.

2.2.0.1 Ultrasound Acquisition Protocol: Data were acquired with subjects in the prone position at three plantar foot locations (hallux, medial sesamoid, and 5th metatarsal head) using a thick gel layer to avoid applying pressure.

2.2.0.2 Plantar Pressure Acquisition Protocol: Plantar pressure was measured using a Zebris FDM force platform during barefoot walking at a self-selected speed. Five trials were recorded (100 Hz sampling rate) after two preliminary trials.

2.2.0.3 Microvascular Status Acquisition Protocol: The Toe Brachial Pressure Index was measured using the MESI mTABLET device.

2.2.0.4 Parameter Estimation From the collected data, ten key parameters were derived:

- Combined thickness of the epidermis and dermis
- Hypodermis thickness
- Radius of curvature of bony prominences, including the hallux, medial sesamoid, and 5th metatarsal head
- Maximum peak plantar pressure obtained from Zebris system
- Stiffness
- Toe-Brachial Index (TBI) measured using MESI device
- Peak Green-Lagrange shear strain calculated from combined datasets

2.2.0.5 Morphological Parameters Ultrasound B-mode images were used to measure skin layer thicknesses and bone curvature at three plantar foot locations: hallux, medial sesamoid, and 5th metatarsal head. The combined epidermis and dermis thickness, as well as hypodermis thickness, were estimated directly on the ultrasound device. Additionally, a custom MATLAB routine computed the radius of curvature of bony prominences by selecting 10 points on the bone surface and applying a least-squares regression.(Figure 1)

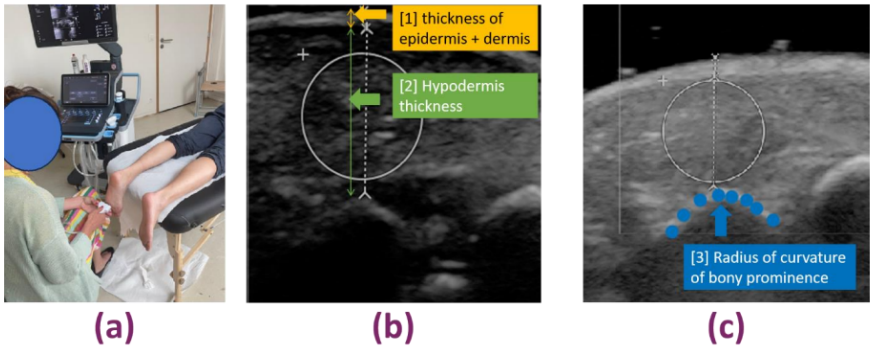


Figure 1: (a) Acquisition in prone position; (b) Ultrasound B-mode image showing skin layers; (c) Radius of curvature estimation from bone surface points.

2.2.0.6 Boundary Conditions (Peak Interface Pressure) In this study, the "seven-zone analysis" proposed by **Zebris** was utilized to automatically segment plantar pressure measurements (Figure 2(b)) into **seven distinct zones** (Figure 2(c)).

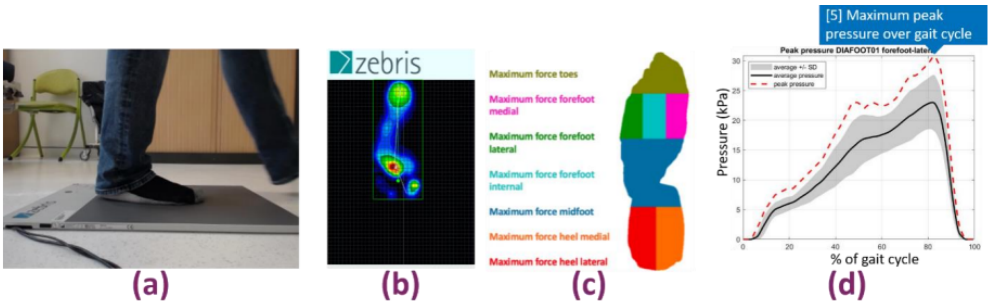


Figure 2: Plantar pressure measurements and analysis. (a) A Zebris force platform used for data collection. (b) Example of plantar pressure distribution. (c) Seven-zone analysis of plantar pressure. (d) Plantar pressure (mean, standard deviation, and peak) over the gait cycle.

Subsequently, the **plantar pressure** (mean, standard deviation, and peak) was plotted (Figure 11(d)) for three regions of interest: the **hallux**, the **medial sesamoid**, and the **5th metatarsal head**. These correspond, respectively, to the toes, the forefoot lateral, and the internal forefoot. The **maximum peak pressure** over the gait cycle (Figure 3 was also extracted.

Figure 11(a) illustrates the use of a **Zebris force platform** to measure plantar pressure parameters during barefoot walking at a self-selected speed. These measurements were conducted over 5 trials at a sampling rate of 100 Hz, with 2 preliminary trials. Figure 3 provides an example of the plantar pressure distribution.

2.3 Available Data

The data utilized for this study is primarily sourced from an **Excel file**, which is expected to contain a dedicated sheet titled "DIAFOOT". This file serves as the central repository for all collected patient-specific measurements and clinical assessments.

The dataset is comprehensive, encompassing various types of data crucial for a holistic assessment of diabetic foot ulcer risk:

- **Clinical Data:** This includes anthropometric measurements such as Height (m) and Weight (kg).
- **Vascular Data:** Obtained using the MESI device, this category includes parameters like Ankle Pressure (Right/Left) and the Ankle Brachial Index (ABI) or Big Toe Systolic Pressure Index (Right/Left).
- **Neuropathic/Sudomotor Function Data:** Measurements from the SUDOSCAN device provide insights into sudomotor function, including SUDOSCAN Hand (Right/Left) and SUDOSCAN Foot (Right/Left) values.
- **Biomechanical Data:** This is a rich category covering several key aspects of foot mechanics:
  - **Range of Motion (ROM):** Including ROM for MTP1 (Right/Left) and Ankle (Right/Left).
  - **Plantar Pressure Analysis:** Average maximum pressures at specific anatomical regions such as SESA, Hallux, and TM5 (Right/Left).
  - **Tissue Stiffness (Duro):** Measured using a durometer at SESA, Hallux, and TM5 (Right/Left), providing insights into soft tissue mechanical properties.
  - **Total Tissue Thickness (Ultrasound):** Ultrasound-derived measurements of total tissue thickness at SESA, Hallux, and TM5 (Right/Left).
  - **Rate of Change (ROC):** Associated with tissue properties at SESA, Hallux, and TM5 (Right/Left).
- **Risk Stratification Data:** The crucial International Working Group on the Diabetic Foot (IWGDF) risk grade (0, 1, 2, or 3) for each patient, serving as the target variable for classification and clustering analyses.

The Excel file is structured such that each column from column 1 onwards represents an individual patient, with rows corresponding to specific parameters. This comprehensive data collection allows for a multi-faceted assessment of DFU risk.

1	Numero d'inclusion DIAFOOT	1	2	3	4
7	Date de naissance jjmmaa	10/17/1960	3/18/1961	2/11/1958	1/13/1974
8	Genre H/F/I	H	H	F	F
9	Téléphone	06 52 89 40 45	06 18 95 75 36	01 75 29 45 54	06 86 21 65 6
10	Code IPP	8009165384	8009445839	8009101033	8001808274
11	Profession	Retraite taxi	Taxi	ite administra	Sans
12	Encadrement social/ familial (Y/N)	Y	N	N	Y
13	Age du diabète (années)	24	24	10	12
14	Type de diabète 1/2/AT	2	2	2	2
15	Insuline (Y/N)	N	Y	Y	Y
16	Hba1c	6.9	10.4	7.8	7.1
17	Grade de risque IWGDF	0	2	0	0
18	Taille (m)	1.63	1.8	1.63	1.57
19	Poids (kg)	59	97	93	75
20	BMI (poids[kg] / taille[m]2)	22.21	29.94	35.00	30.43
21	By-pass (Y/N)	N	N	N	N
22	Tabac 0/1sevré>3ans/2actif ou sevré<3ans	0	0	0	N

Figure 3: Example of data structure for the DFU study

3 Methodology

The statistical analysis workflow, developed during the internship, processes and analyzes biomechanical and clinical data from the dataset to evaluate diabetic foot health risks. Implemented in Python, the workflow is integrated into an interactive dashboard built with the **Streamlit** framework, facilitating dynamic data exploration and visualization. The dashboard is accessible at:

https://diafoot-analysis.streamlit.app/

The workflow leverages a structured Excel file containing clinical and biomechanical data within a dedicated `DIAFOOT` sheet. This organization supports a comprehensive, multi-parametric analysis of diabetic foot ulcer (DFU) risk.
Users can upload the Excel dataset at the following link: GitHub Excel File through the interface, which automatically triggers the data parsing, cleaning, statistical analysis, and result visualization. The interface offers multiple analysis modules, including:

- **Descriptive and comparative statistics**, normality tests, and hypothesis testing.
- **Biomechanical symmetry analysis** (e.g., Hallux/SESA/TM5 zones, left vs. right).
- **Correlation analysis**, examining relationships between key biomechanical and clinical parameters (e.g., plantar pressure, stiffness, tissue thickness) using Pearson or Spearman methods, with visualizations like scatter plots, regression lines, pair plots, and correlograms.
- **Diabetic vs. control group comparisons**, analyzing differences in biomechanical and clinical parameters (e.g., total tissue thickness, plantar pressure, stiffness) between diabetic neuropathy and control groups based on IWGDF classification.
- **Clustering analysis**, applying KMeans, Agglomerative, and Gaussian Mixture Models to group patients based on biomechanical and clinical features, with evaluation metrics (e.g., Silhouette, ARI) and comparison to IWGDF risk grades.
- **Bland-Altman analysis**, assessing agreement between left and right foot measurements for parameters like hardness, epidermis/dermis thickness, and hypodermis thickness, with pooled plots across anatomical zones (Hallux, SESA, TM5).

The dashboard dynamically loads the appropriate parameter mappings and analysis logic depending on the selected module. All processing and visualization are performed using the following stack:
Streamlit for web deployment and interface rendering.

3.1 Statistical Tools and Software Used

All data processing, statistical analyses, and machine learning model development were conducted using **Python** within an interactive **Streamlit** framework, configured with a wide layout and titled "DIAFOOT Analysis Dashboard" for dynamic data visualization and user interaction.

The following Python libraries and tools were primarily employed:

- **pandas** and **numpy** for data manipulation, cleaning, and numerical computations.
- **scipy.stats** for statistical hypothesis tests, including **ttest_ind**, **ttest_rel**, **mannwhitneyu**, **wilcoxon**, **shapiro**, **chi2_contingency**, **fisher_exact**, and **f_oneway**.
- **statsmodels** for advanced statistical modeling, including Multivariate Analysis of Variance (MANOVA) and regression modeling.
- **scikit-learn** for machine learning workflows, including data preprocessing (e.g., **StandardScaler**, **SimpleImputer**), clustering (e.g., **KMeans**, **AgglomerativeClustering**, **GaussianMixture**), dimensionality

reduction (e.g., **PCA**), regression (e.g., **LinearRegression**, **LogisticRegression**), classification (e.g., **RandomForestClassifier**), and evaluation metrics such as **silhouette_score**, **adjusted_rand_score**, **normalized_mutual_info_score**, **calinski_harabasz_score**, **davies_bouldin_score**, and **confusion_matrix**.

- **matplotlib** and **seaborn** for comprehensive data visualization, including heatmaps, scatterplots, distribution plots, and confusion matrix displays, with support from **matplotlib.cm** for colormap handling.
- Python's built-in modules like **datetime**, **io**, **re**, and **io.BytesIO** for handling temporal data, file input/output operations, and regular expressions.
- **streamlit** for developing an interactive web-based dashboard for clinical decision support.

These tools collectively enabled robust exploratory data analysis, hypothesis testing, predictive modeling, clustering, and the development of an interactive dashboard for clinical decision support.

4 Statistical Analysis

The statistical analysis workflow developed during the internship focused on processing and analyzing biomechanical and clinical data from the DIAFOOT dataset to assess diabetic foot health risks. The workflow was implemented using Python and integrated into a Streamlit-based dashboard, enabling interactive data exploration and visualization. Below is an overview of the key steps in the workflow:

4.1 Data Ingestion and Preprocessing

- **File Upload**: The workflow begins with uploading an Excel file containing the DIAFOOT sheet, which includes biomechanical and clinical parameters such as plantar pressure, tissue stiffness, ultrasound thickness, and IWGDF risk grades.
- **Data Cleaning**: The dataset is cleaned by extracting relevant rows (e.g., BMI, MESI ankle pressure, Michigan Score) and converting them to numeric formats. Missing values are handled using mean imputation, and non-numeric entries (e.g., 'NR') are replaced with NaN for consistency.
- **Feature Selection**: A predefined set of target parameters is selected based on their clinical relevance, such as tissue thickness, stiffness, and temperature measurements across anatomical zones (Hallux, SESA, TM5).
- **Data Transformation**: The data is transposed to represent patients as rows and features as columns. Additional features, such as age (derived from date of birth) and AOMI (extracted from strings), are computed to enrich the dataset.

4.2 Exploratory Data Analysis (EDA)

- **Basic Analysis**: Summary statistics and visualizations (e.g., histograms for age and BMI, pie charts for gender, bar plots for IWGDF risk grades) are generated to understand patient demographics and risk distributions.
- **Descriptive Analysis**: For each parameter, summary statistics (mean, median, standard deviation, quartiles) are calculated, and normality is tested using the Shapiro-Wilk test. Boxplots visualize parameter distributions.

- **Correlation Analysis**: A correlation matrix is computed to identify relationships between parameters, with high correlations ($|r| \geq 0.8$) highlighted in a downloadable Excel report. Visualizations include scatter plots, regression plots, pair plots, and correlograms.

4.3 Comparative Statistical Analysis

- **Left vs. Right Foot Comparison**: Paired t-tests or Wilcoxon signed-rank tests (based on normality) compare left and right foot parameters (e.g., stiffness, pressure) across zones. Results are visualized with boxplots and exported as Excel files.
- **Diabetic vs. Control Analysis**: Parameters are pooled across left and right feet and compared between diabetic neuropathy (IWGDF ≥ 1) and control (IWGDF = 0) groups. Summary statistics (mean, standard deviation, range) are presented in tables and boxplots.
- **Risk Grade Transitions**: Statistical tests (t-tests or Mann-Whitney U tests) compare parameters between successive IWGDF grades (0 vs. 1, 1 vs. 2, 2 vs. 3) and combined groups (Grades 0-1 vs. 2-3). Significant differences ($p < 0.05$) are reported.

4.4 Clustering and Multivariate Analysis

Clustering analyses were conducted both with and without the inclusion of IWGDF grades to investigate intrinsic patient groupings based on the full set of biomechanical and clinical parameters. Four unsupervised algorithms were applied:

- **Agglomerative (Ward)**: Hierarchically merges clusters by minimizing within-cluster variance.
- **Agglomerative (Average)**: Hierarchically merges clusters using the average inter-point distance.
- **KMeans**: Partitions the data into k clusters by minimizing the distance to cluster centroids.
- **Gaussian Mixture Model (GMM)**: Clusters data using probabilistic Gaussian distributions.

The number of clusters (k) was user-defined, ranging from 2 to 6 (default: 4).

4.4.0.1 Parameter Selection and Preprocessing Clustering analyses were conducted both with and without the inclusion of IWGDF grades to investigate intrinsic patient groupings. Each analysis was performed twice:

1. **Using only the most important parameters:** IWGDF grade, MESI index, Michigan score, plantar pressure, tissue stiffness, tissue thickness, amplitude, and temperature.
2. **Using the complete feature set:** Over 70 variables spanning anthropometric, clinical, and biomechanical domains.

Categorical variables (e.g., Charcot history) were one-hot encoded, features with excessive missingness were excluded, and remaining missing values were imputed using the median. All features were standardized prior to clustering.

4.4.0.2 Dimensionality Reduction and Clustering Principal Component Analysis (PCA) was applied to reduce data dimensionality and allow 2D visualization. Clustering performance was evaluated using the following metrics:

- **Silhouette Score**: Measures cluster cohesion and separation (−1 to 1; higher is better).
- **Calinski–Harabasz Score**: Quantifies between- vs. within-cluster dispersion (higher is better).
- **Davies–Bouldin Score**: Evaluates cluster similarity (lower is better).
- **Adjusted Rand Index (ARI)**: Compares predicted clusters to true IWGDF grades (higher is better).
- **Normalized Mutual Information (NMI)**: Measures information overlap between predicted and true labels (higher is better).

Confusion matrices were generated to compare cluster assignments with IWGDF grades.

4.4.0.3 Feature Importance and Group Differences Feature importance was determined using three complementary approaches:

- **Logistic Regression with LASSO regularization**
- **Random Forest Classifier**
- **Univariate ANOVAs** (with eta-squared effect sizes)

MANOVA was then used to evaluate differences between IWGDF groups, with p-values indicating significant separations in the parameter space.

4.5 Correlation

A correlation matrix is computed for key biomechanical, vascular, thermographic, and clinical parameters. Highly correlated pairs (|| 0.8 by default) are extracted and visualized using:

- *Scatter plots* and *regression lines* to highlight linear trends.
- *Pair plots (scatterplot matrices)* for multivariate relationships.
- *Correlograms* (heatmaps) showing only strong correlations (|| 0.7).

Results are exportable and help identify redundancies or hidden associations between parameters.

5 Web Dashboard Development (Streamlit)

- **Interactive Dashboard**: The Streamlit dashboard allows users to select analysis types (e.g., Basic Analysis, Normality Tests, Clustering) and visualize results through histograms, boxplots, scatter plots, and heatmaps.
- **Exportable Outputs**: Results, including summary statistics, test results, and clustering metrics, are exported as Excel files for further analysis and reporting.

This workflow provides a comprehensive framework for analyzing DIAFOOT data, enabling the identification of biomechanical and clinical differences associated with diabetic foot risk. It supports clinical decision-making by highlighting significant parameters and visualizing patient groupings.

6 Results

The results presented in this section aim to characterize the biomechanical and clinical profiles of the DIAFOOT dataset cohort and identify patterns associated with diabetic foot ulcer (DFU) risk. Each analysis—descriptive statistics, normality tests, group comparisons, clustering, and correlations—was conducted to uncover relationships between demographic, clinical, and biomechanical variables and DFU risk, as defined by the International Working Group on the Diabetic Foot (IWGDF) classification. The findings build a foundation for understanding how these variables contribute to risk stratification and inform subsequent predictive modeling.

6.1 Descriptive Statistics and Population Characteristics

Descriptive statistics provide an initial overview of the cohort's demographic and clinical characteristics, setting the stage for deeper analyses by highlighting variability and potential risk factors for DFU.

6.1.0.1 Age Distribution The cohort's mean age was 56.8 (SD = 14.1), reflecting a broad age range that captures both early-onset and long-standing diabetes cases. This variability suggests diverse disease progression stages, which may influence DFU risk.

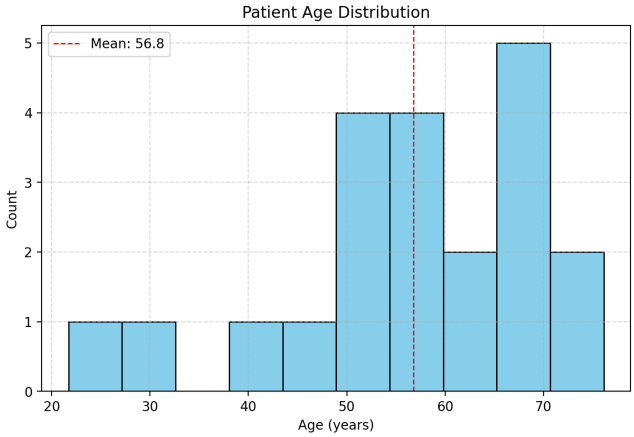


Figure 4: Histogram of patient age distribution (mean = 56.8, SD = 14.1).

6.1.0.2 Gender Distribution The cohort exhibited a nearly balanced gender distribution, ensuring that findings are not biased by sex and are broadly applicable to diabetic populations.

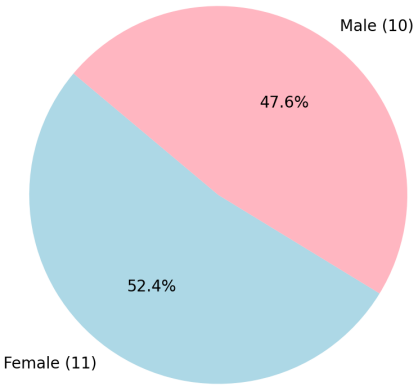


Figure 5: Gender distribution of the cohort.

6.1.0.3 Body Mass Index (BMI) BMI, a known contributor to foot loading and DFU risk, averaged 25.83 (kg/m^2), indicating a high prevalence of overweight individuals, consistent with literature linking elevated BMI to increased plantar pressure.

Table 1: Summary statistics for BMI

Statistic	Min	Q1	Median	Mean	Q3	Max
BMI (kg/m^2)	18.13	21.65	25.49	25.83	30.3	35

6.1.0.4 IWGDF Risk Grade Distribution Patients were classified using the IWGDF risk grades (0–3), with a polarized distribution: 11 patients at Grade 0 (low risk), 2 at Grade 1, 1 at Grade 2, and 7 at Grade 3 (very high risk). This skew highlights distinct risk profiles, with most patients either at minimal or severe risk, necessitating targeted analyses to differentiate these groups.

6.1.0.5 Age of Diabetes by Type The age at diabetes diagnosis varied by type, with Type 1, Type 2, and atypical forms showing distinct patterns. This differentiation helps contextualize disease duration as a potential DFU risk factor.

6.1.0.6 Age of Diabetes by Type: This chart illustrates the age at diagnosis (i.e., "age of diabetes") across different types of diabetes:

beginitemize

■ **Type 1**

■ **Type 2**

■ **Atypical (AT)**: Other or rare forms of diabetes.

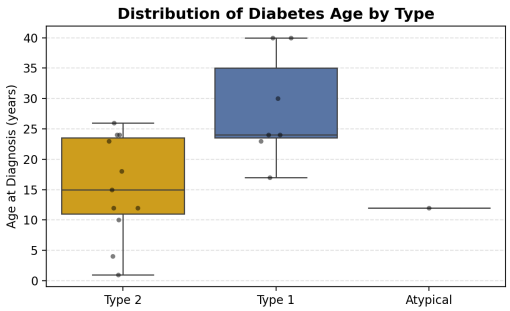


Figure 6: Age at Diagnosis by Diabetes Type

6.2 Normality and Distribution of Key Parameters

Normality testing was conducted to guide the selection of appropriate statistical methods for subsequent analyses, ensuring robust comparisons and clustering. The Shapiro–Wilk test was applied to key biomechanical and clinical parameters.

The table below summarizes the key descriptive statistics (mean, median, standard deviation, quartiles, min/max) and the results of the Shapiro–Wilk test for normality for selected clinical and biomechanical parameters. Parameters with p-value > 0.05 were considered approximately normally distributed.

Label	Mean	Median	SD	Min	Q1	Q3	Max	W	p-value	Normal
Height (m)	1.62	1.69	0.37	0.09	1.63	1.75	1.92	0.4588	0.0000	No
Weight (kg)	71.96	71.92	20.86	17.37	59.75	86.25	104	0.9447	0.2933	Yes
BMI	25.83	25.49	5.29	18.13	21.65	30.30	35.00	0.9473	0.3836	Yes
MESI Ankle Pressure R	121.41	131.62	42.13	39.17	96.50	150.25	182.00	0.9538	0.4878	Yes
MESI Ankle Pressure L	130.70	139.00	41.72	36.22	101.00	155.00	188.00	0.9264	0.1896	Yes
MESI Big Toe Index R	0.97	1.06	0.31	0.27	0.86	1.17	1.45	0.9162	0.1107	Yes
MESI Big Toe Index L	1.04	1.09	0.30	0.22	0.87	1.28	1.32	0.8425	0.0082	No
Michigan Score	3.25	0.00	5.18	0.00	0.00	8.00	13.00	0.6220	0.0000	No
Amplitude MTP1 R	75.93	80.00	19.54	12.01	73.50	85.75	102.00	0.7900	0.0011	No
Amplitude Talo-crurale L	18.14	18.00	11.60	-15.00	14.11	23.50	43.00	0.9154	0.0932	Yes
Avg Pressure Max SESA R	32.71	24.39	23.09	9.00	19.00	35.09	96.00	0.7664	0.0004	No
Stiffness HALLUX R	15.28	15.32	9.31	1.50	9.66	19.50	38.00	0.9565	0.5992	Yes
US Épaisseur ED HALLUX R	0.94	0.97	0.32	0.29	0.70	1.11	1.58	0.9743	0.8735	Yes
US Épaisseur Hypoderme TM5 R	9.12	9.17	2.39	1.65	8.41	10.15	14.05	0.8509	0.0069	No
Total Thickness TM5 L	1.42	0.00	3.52	0.00	0.00	0.00	13.72	0.4436	0.0000	No
ROC HALLUX R	67.92	58.71	34.50	28.41	37.60	102.33	130.64	0.8843	0.0309	No
Temperature Heel R	25.13	26.90	6.25	2.07	24.00	27.70	29.80	0.5534	0.0000	No
Average IR Temp. Foot R	26.57	26.45	1.79	24.29	25.38	27.04	30.66	0.9261	0.2380	Yes
SUDOSCAN Hand R	42.04	46.00	19.96	11.00	22.50	51.50	81.00	0.9401	0.2905	Yes
SUDOSCAN Foot L	51.55	56.62	22.42	19.00	26.75	71.25	78.00	0.8590	0.0118	No

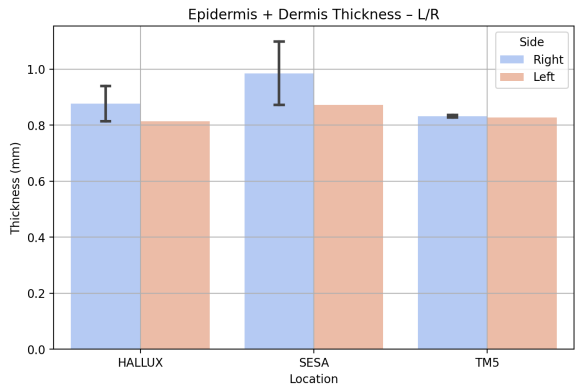
Table 2: Descriptive statistics and Shapiro–Wilk normality test results for selected parameters.

Most parameters, such as BMI and MESI ankle pressures, were normally distributed ($p > 0.05$), allowing parametric tests like t-tests. Non-normal parameters, such as Michigan Score and plantar pressure, required non-parametric tests (e.g., Wilcoxon signed-rank), ensuring robust statistical comparisons.

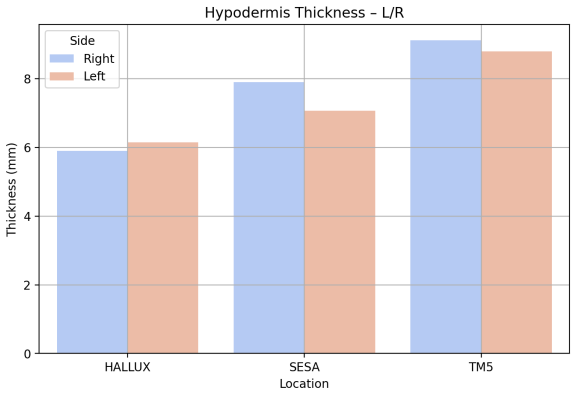
6.3 Comparison by Anatomical Zone

To investigate biomechanical asymmetries that may contribute to DFU risk, parameters such as tissue thickness, stiffness, and pressure were compared across three anatomical zones (Hallux, SESA, TM5) for both feet. This analysis tests the hypothesis that asymmetrical loading or tissue properties could indicate early risk factors.

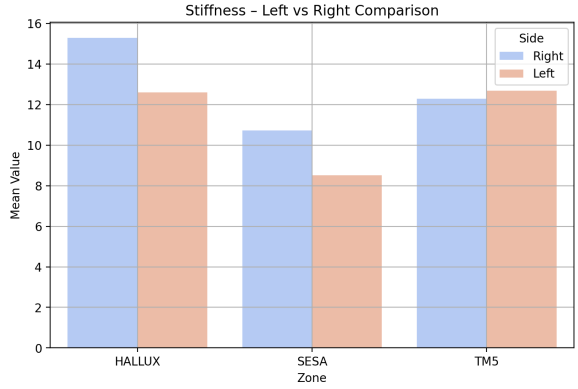
A bar chart is provided separately (see Figure 7).



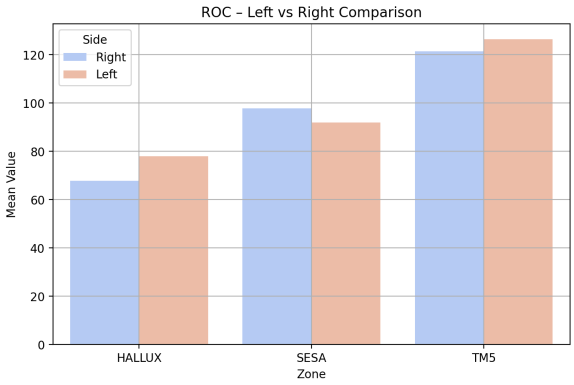
(a) Combined Epidermis + Dermis Thickness: Left vs Right Comparison across Hallux, SESA, and TM5 zones.



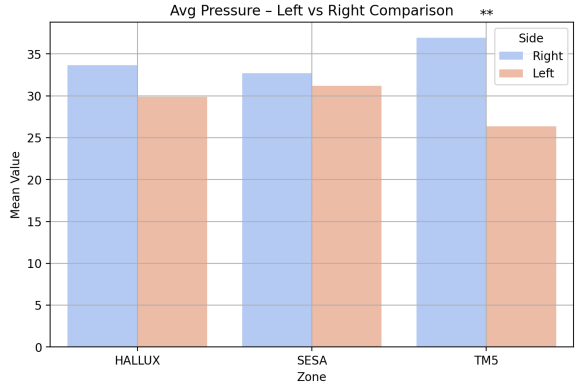
(b) Hypodermis Thickness: Left vs Right Comparison across Hallux, SESA, and TM5 zones.



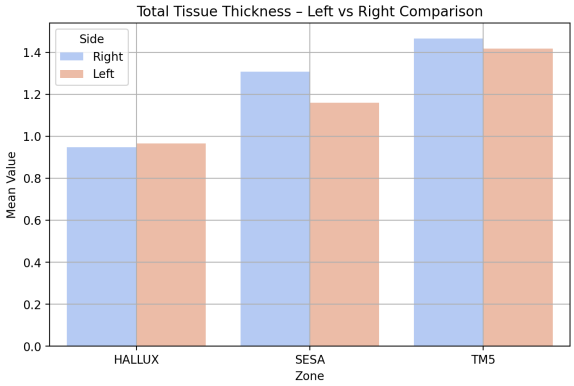
(c) Stiffness: Left vs Right Comparison across Hallux, SESA, and TM5 zones.



(d) ROC (Radius of Curvature): Left vs Right Comparison across Hallux, SESA, and TM5 zones.



(e) Average Pressure: Left vs Right Comparison across Hallux, SESA, and TM5 zones.



(f) Total Tissue Thickness: Left vs Right Comparison across Hallux, SESA, and TM5 zones.

Figure 7: Comparison of various parameters between Left and Right feet across different zones.

Zone	Side	Mean (mm)	SD (mm)
Hallux	Right	0.94	0.32
Hallux	Left	0.815	0.243
Hallux	p-value	NaN	–
SESA	Right	1.099	0.501
SESA	Left	0.872	0.324
SESA	p-value	NaN	–
TM5	Right	0.837	0.395
TM5	Left	0.828	0.281
TM5	p-value	NaN	–

(a) Epidermis + Dermis Thickness (identical values).

Zone	Side	Mean (mm)	SD (mm)
Hallux	Right	5.891	1.678
Hallux	Left	6.146	1.815
Hallux	p-value	0.4486	–
SESA	Right	7.901	3.721
SESA	Left	7.063	2.644
SESA	p-value	0.3325	–
TM5	Right	9.115	2.386
TM5	Left	8.798	2.297
TM5	p-value	0.2242	–

(b) Hypodermis Thickness.

Zone	Side	Mean (N/mm)	SD (N/mm)
Hallux	Right	15.283	9.314
Hallux	Left	12.592	8.083
Hallux	p-value	0.0535	–
SESA	Right	10.730	10.996
SESA	Left	8.522	5.353
SESA	p-value	0.3962	–
TM5	Right	12.280	9.217
TM5	Left	12.673	9.576
TM5	p-value	0.8412	–

(c) Stiffness.

Zone	Side	Mean (mm)	SD (mm)
Hallux	Right	67.920	34.498
Hallux	Left	77.929	35.599
Hallux	p-value	0.1857	–
SESA	Right	97.910	46.518
SESA	Left	91.921	51.034
SESA	p-value	0.3473	–
TM5	Right	121.445	55.355
TM5	Left	126.446	46.469
TM5	p-value	0.6513	–

(d) Radius of Curvature (ROC).

Zone	Side	Mean (kPa)	SD (kPa)
Hallux	Right	33.638	13.773
Hallux	Left	29.891	11.387
Hallux	p-value	0.3035	–
SESA	Right	32.714	23.086
SESA	Left	31.183	15.035
SESA	p-value	0.6048	–
TM5	Right	36.929	19.092
TM5	Left	26.338	12.984
TM5	p-value	**0.0054**	–

(e) Average Pressure.

Zone	Side	Mean (mm)	SD (mm)
Hallux	Right	0.949	2.444
Hallux	Left	0.966	2.497
Hallux	p-value	0.8562	–
SESA	Right	1.308	3.471
SESA	Left	1.16	2.973
SESA	p-value	0.2258	–
TM5	Right	1.466	3.634
TM5	Left	1.419	3.515
TM5	p-value	0.2025	–

(f) Total Tissue Thickness.

Figure 8: Comparison of right and left foot parameters across Hallux, SESA, and TM5 zones.

Most parameters showed no significant bilateral differences ($p > 0.05$), suggesting general symmetry in tissue properties. However, a significant asymmetry in average pressure at the TM5 zone ($p = 0.0054$; right: 36.93 kPa, left: 26.34 kPa) indicates potential uneven loading or compensatory gait patterns, which could elevate DFU risk on the right foot. This finding prompts further investigation into lateral forefoot biomechanics. Other parameters, such as epidermis and dermis thickness, hypodermis thickness, stiffness, and total tissue thickness, show consistent patterns between sides, with no statistically meaningful discrep-

ancies. Notably, some comparisons yielded non-numeric p-values (NaN), indicating either insufficient data or non-applicable statistical tests.

6.4 Paired Comparison of Left and Right Foot Parameters

Paired comparisons of bilateral parameters were performed to quantify asymmetries that could signal early biomechanical or vascular changes relevant to DFU risk. Appropriate statistical tests (t-test or Wilcoxon) were selected based on normality results.

Table 3: Statistical Comparison of Left and Right Foot Parameters

Parameter Pair	Test Used	Statistic	p-value	Significant
MESI Ankle Pressure R vs L	Paired t-test	-0.3696	0.7165	No
MESI Big Toe Systolic Pressure Index R vs L	Wilcoxon signed-rank test	60.5000	0.6981	No
Medical history of acute Charcot R vs L	–	–	–	Not enough data (n=0)
Amplitude of dorsiflexion of right MTP1 R vs L	Wilcoxon signed-rank test	63.5000	0.5377	No
Amplitude talo-crurale R vs L	Paired t-test	0.5273	0.6044	No
Avg Pressure Max SESA R vs L	Wilcoxon signed-rank test	84.0000	0.6794	No
Avg Pressure Max HALLUX R vs L	Paired t-test	1.0610	0.3035	No
Avg Pressure Max TM5 R vs L	Wilcoxon signed-rank test	19.0000	0.0038	Yes
Stiffness SESA R vs L	Wilcoxon signed-rank test	58.0000	0.9095	No
Stiffness HALLUX R vs L	Wilcoxon signed-rank test	25.5000	0.0498	Yes
Stiffness TM5 R vs L	Paired t-test	-0.2036	0.8412	No
US Épaisseur ED SESA R vs L	Paired t-test	1.7663	0.0943	No
US Épaisseur ED HALLUX R vs L	Wilcoxon signed-rank test	53.0000	0.1674	No
US Épaisseur ED TM5 R vs L	Paired t-test	0.0849	0.9332	No
US Épaisseur Hypoderme SESA R vs L	Wilcoxon signed-rank test	57.0000	0.1336	No
US Épaisseur Hypoderme HALLUX R vs L	Paired t-test	-0.7757	0.4486	No
US Épaisseur Hypoderme TM5 R vs L	Wilcoxon signed-rank test	76.0000	0.4653	No
Total Tissue Thickness SESA R vs L	Wilcoxon signed-rank test	54.0000	0.0990	No
Total Tissue Thickness HALLUX R vs L	Wilcoxon signed-rank test	87.0000	0.7475	No
Total Tissue Thickness TM5 R vs L	Wilcoxon signed-rank test	75.0000	0.4209	No
ROC SESA R vs L	Wilcoxon signed-rank test	73.0000	0.3955	No
ROC HALLUX R vs L	Wilcoxon signed-rank test	49.0000	0.1187	No
ROC TM5 R vs L	Paired t-test	-0.4596	0.6513	No
Temperature Hallux R vs L	Wilcoxon signed-rank test	55.0000	0.5282	No
Temperature 5th Toe R vs L	Wilcoxon signed-rank test	58.5000	0.6685	No
Temperature Plantar Arch R vs L	Wilcoxon signed-rank test	21.0000	0.0067	Yes
Temperature Lateral Sole R vs L	Wilcoxon signed-rank test	11.5000	0.0011	Yes
Temperature Forefoot R vs L	Wilcoxon signed-rank test	42.5000	0.1871	No
Temperature Heel R vs L	Wilcoxon signed-rank test	76.0000	1.0000	No
Temperature Difference Hand-Foot R vs L	Wilcoxon signed-rank test	47.5000	0.4887	No
Normalized Temperature R vs L	Wilcoxon signed-rank test	40.0000	0.2769	No
SUDOSCAN Hand R vs L	Paired t-test	-0.3393	0.7385	No
SUDOSCAN Foot R vs L	Wilcoxon signed-rank test	51.0000	0.3742	No

Significant asymmetries were observed in average pressure at TM5 ($p = 0.0038$), stiffness at Hallux ($p = 0.0498$), and temperatures at the plantar arch ($p = 0.0067$) and lateral sole ($p = 0.0011$). Asymmetry in itself does not indicate any particular risk, except for temperature. Temperature asymmetry could be an inflammatory sign for grade 1, 2 or 3 patients (possibly Charcot syndrome) in a given patient.

6.5 Group Comparison: Grades 0–1 vs. 2–3

To assess how biomechanical and clinical parameters differ across DFU risk levels, patients were grouped into low-to-moderate risk (Grades 0–1, n=13) and high-to-very-high risk (Grades 2–3, n=8). This analysis identifies variables that distinguish risk groups, informing targeted interventions.

Several parameters exhibited statistically significant differences ($p < 0.05$) between the two groups:

Table 4: Significant Differences Between IWGDF Groups A (Grades 0–1) and B (Grades 2–3)

Parameter	Test	p-value	Signif.	Group A Mean	Group B Mean
IWGDF Risk Grade	Mann–Whitney U	3.8×10^{-5}	Yes	0.15	2.88
Michigan Score (MNSI)	Mann–Whitney U	0.00053	Yes	11.69	4.13
Height (m)	t-test	0.0095	Yes	1.67	1.79
Total Tissue Thickness SESA R	Mann–Whitney U	0.0099	Yes	9.55	4.17
Temperature Heel R (°C)	t-test	0.0188	Yes	25.85	28.55
Total Tissue Thickness HALLUX R	Mann–Whitney U	0.0315	Yes	7.19	2.59
Temperature Heel L (°C)	t-test	0.0387	Yes	25.76	28.18
SUDOSCAN Foot L (μS)	Mann–Whitney U	0.0393	Yes	59.00	28.33
SUDOSCAN Foot R (μS)	Mann–Whitney U	0.0433	Yes	59.85	26.67
Temperature Forefoot R (°C)	Mann–Whitney U	0.0498	Yes	25.25	27.53

Significant differences in Michigan Score, tissue thickness, temperature, and SUDOSCAN readings between low- and high-risk groups highlight their role in risk stratification. Higher temperatures and lower tissue thickness in the high-risk group suggest vascular and structural changes associated with advanced neuropathy, supporting their use in predictive models. Other parameters analyzed did not show statistically significant inter-group differences.

6.6 Clustering Analysis

Clustering was performed to identify natural patient subgroups based on biomechanical and clinical features, testing whether unsupervised methods can replicate or enhance IWGDF risk stratification. Two approaches were used: one with a focused set of key parameters and another with all 87 variables. The results for each stage are presented below, with comparisons between clustering conducted **with** and **without** the IWGDF grade as an input variable.

6.6.0.1 Clustering with Important Parameters Using key parameters (IWGDF grade, Michigan Score, MESI, pressure, stiffness, thickness, amplitude, temperature), clustering produced compact groups, particularly when including IWGDF grade (Silhouette score = 0.553). This suggests that these parameters capture clinically relevant variability.

6.6.0.1.1 Clustering Comparison: With vs. Without IWGDF Grade Figure 9 shows the PCA-based visualization of clustering results obtained using the most important parameters, both with and without the IWGDF grade as a feature. The clustering structures appear more compact when the IWGDF grade is included, suggesting that this variable provides meaningful separation among patients.

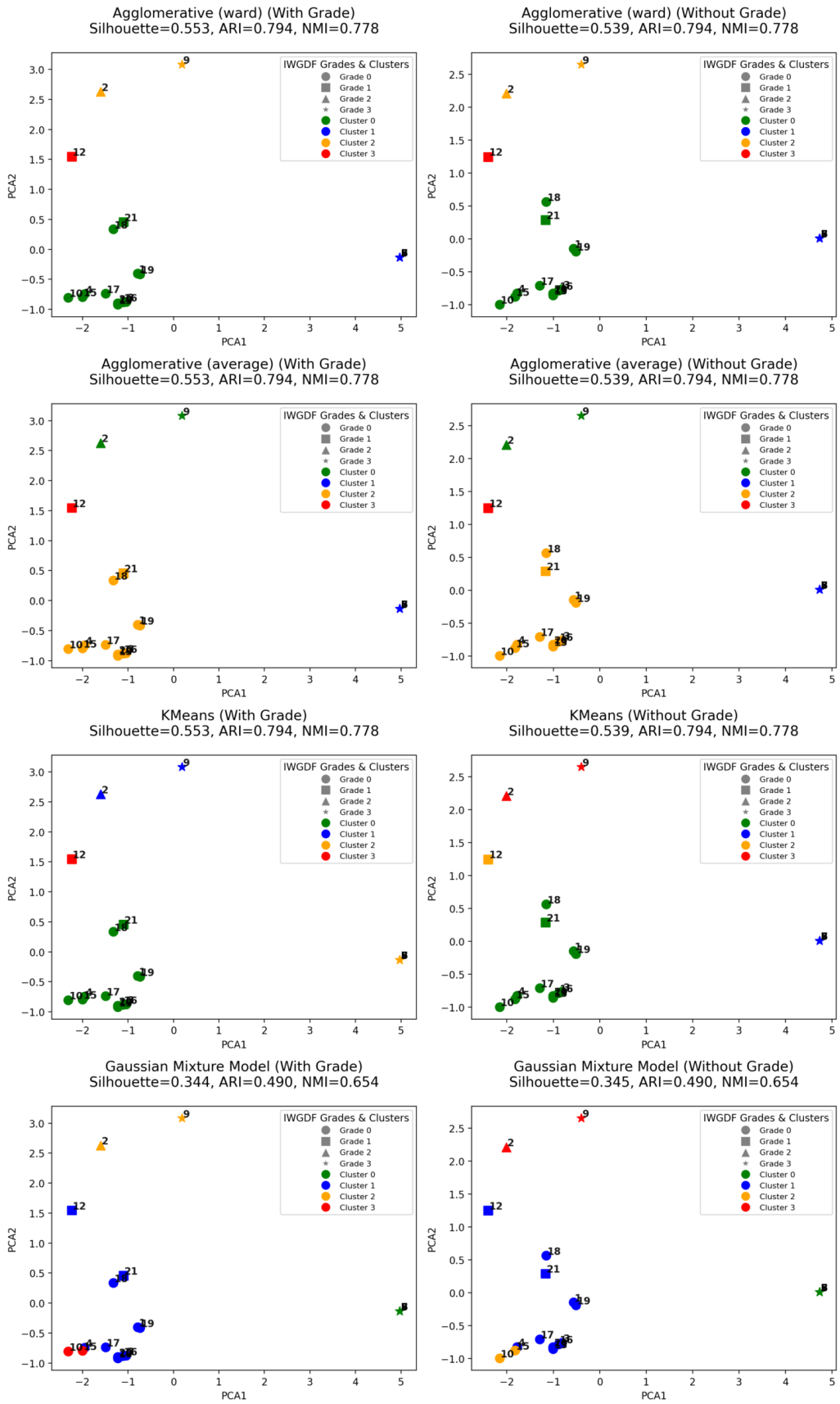


Figure 9: Clustering comparison using the most important parameters, with and without IWGDF grade.

Patients 2(grade 2) and 9 (grade 3) seem fairly close Grade 0s are clearly isolated in their cluster++++ Patient 21 is a grade 1 close to grade 0s

6.6.0.1.2 Clustering Metrics Table 5 summarizes clustering performance metrics for all four algorithms. Including the IWGDF grade led to slightly higher Silhouette and Calinski–Harabasz scores and lower Davies–Bouldin scores, indicating improved separation and compactness of the clusters.

Table 5: Clustering Metrics (With and Without IWGDF Grade) using Important Parameters

Algorithm	Silhouette	CH Score	DB Score	ARI	NMI
With IWGDF Grade					
Agglomerative (Ward)	0.553	43.959	0.485	0.794	0.778
Agglomerative (Average)	0.553	43.959	0.485	0.794	0.778
KMeans	0.553	43.959	0.485	0.794	0.778
Gaussian Mixture Model	0.344	33.367	0.939	0.490	0.654
Without IWGDF Grade					
Agglomerative (Ward)	0.539	40.721	0.532	0.794	0.778
Agglomerative (Average)	0.539	40.721	0.532	0.794	0.778
KMeans	0.539	40.721	0.532	0.794	0.778
Gaussian Mixture Model	0.345	31.163	0.957	0.490	0.654

6.6.0.1.3 Homogeneity and Sample Distribution The within-group inertia for the IWGDF groups was 43.875, indicating moderate variability within grades. Sample sizes per grade were as follows: Grade 0: 11, Grade 1: 2, Grade 2: 1, Grade 3: 5.

6.6.0.1.4 Feature Importance Feature importance analyses revealed the following:

- **Logistic Regression with LASSO:** Michigan Score (1.94), Michigan Score 2 (0.94), and Total Tissue Thickness (SESA_R) (0.11) had the strongest predictive contributions.
- **Random Forest:** Total Tissue Thickness (TM5_L, 0.161) and (SESA_R, 0.160) were top predictors, followed by Michigan Score (0.157).
- **Univariate ANOVA (Eta Squared):** Total Tissue Thickness (HALLUX_R, 0.827) and Michigan Score (0.827) showed the highest effect sizes.

Feature importance rankings for each clustering algorithm are summarized below:

- **With IWGDF Grade:** Michigan Score and IWGDF grade consistently ranked highest, followed by Total Tissue Thickness (TM5_L and SESA_R).
- **Without IWGDF Grade:** Michigan Score and Total Tissue Thickness (SESA_R, HALLUX_R, TM5_L) dominated the rankings across methods.

6.6.0.2 Clustering with All Parameters Using all 87 parameters resulted in less distinct clusters (Silhouette score = 0.18–0.22), likely due to noise from less-informative features. This underscores the importance of parameter selection for robust clustering.

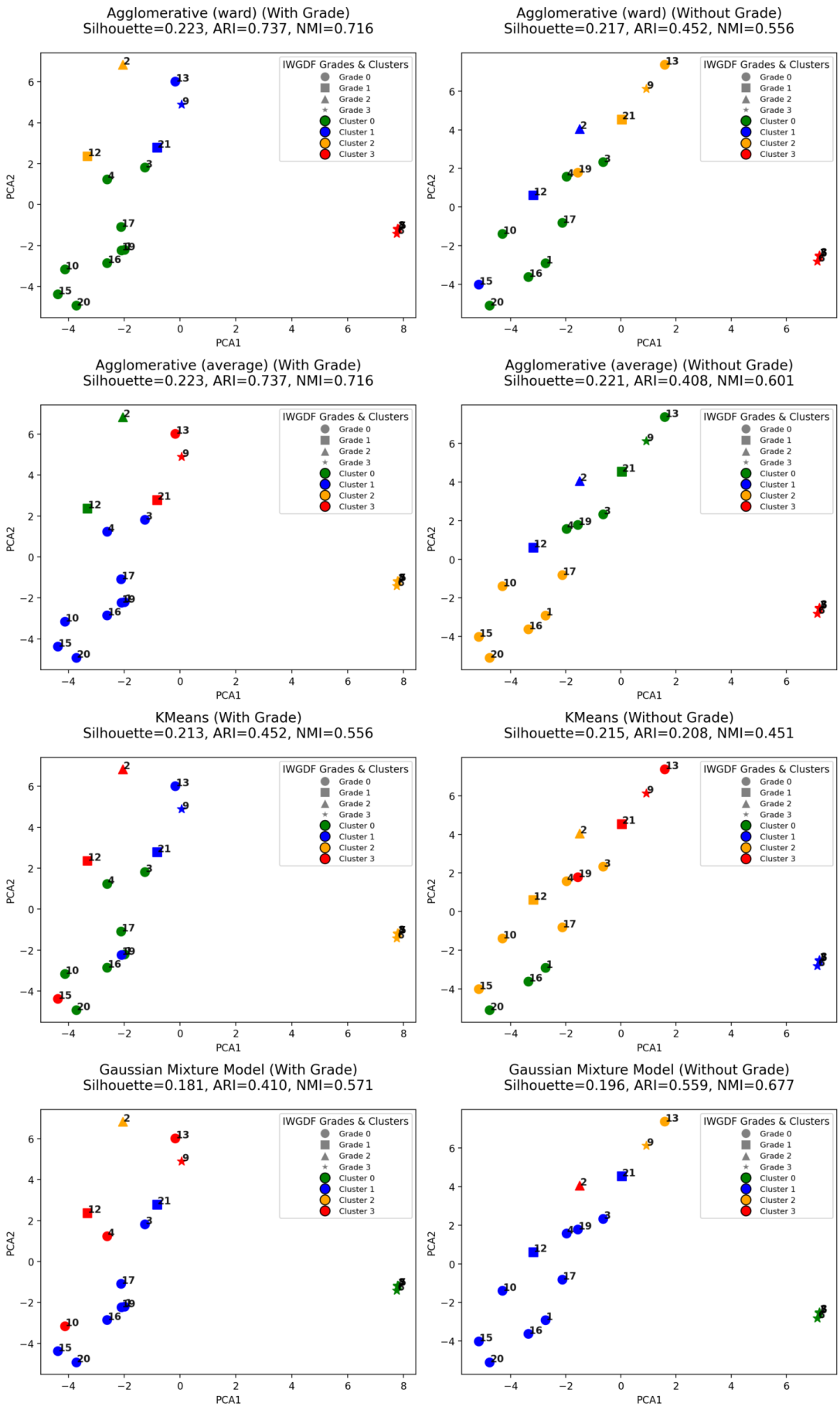


Figure 10: Clustering comparison using all parameters, with and without IWGDF grade.

Table 6: Clustering Metrics (With and Without IWGDF Grade) using All Parameters

Algorithm	Silhouette	CH Score	DB Score	ARI	NMI
With IWGDF Grade					
Agglomerative (Ward)	0.223	3.585	1.498	0.737	0.716
Agglomerative (Average)	0.223	3.585	1.498	0.737	0.716
KMeans	0.213	3.480	1.689	0.452	0.556
Gaussian Mixture Model	0.181	2.962	1.769	0.410	0.571
Without IWGDF Grade					
Agglomerative (Ward)	0.217	3.550	1.639	0.452	0.556
Agglomerative (Average)	0.221	3.661	1.559	0.408	0.601
KMeans	0.215	3.554	1.676	0.208	0.451
Gaussian Mixture Model	0.196	3.050	1.204	0.559	0.677

6.6.0.2.1 Clustering Comparison: With vs. Without IWGDF Grade

6.6.0.2.2 Confusion Matrices Figure 11 displays confusion matrices comparing IWGDF grades and clusters.

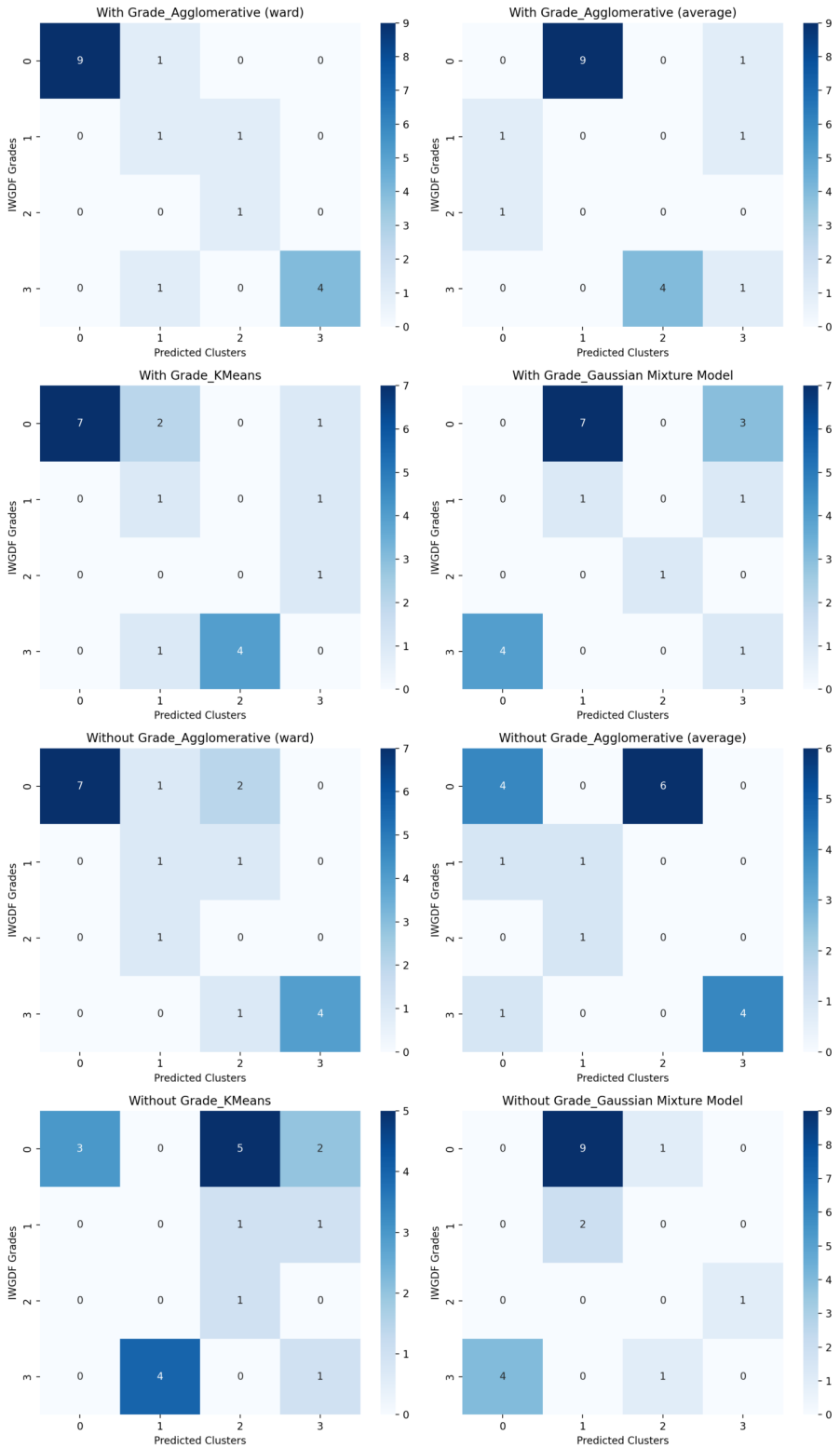


Figure 11: Confusion matrices: all IWGDF grades vs. cluster assignments.

6.6.0.2.3 Homogeneity and Sample Distribution The within-group inertia for IWGDF groups was 981.748, reflecting substantial heterogeneity within grades.

6.6.0.2.4 Feature Importance Top predictive features identified across methods included:

- **LASSO:** Height (1.67), Michigan Score (1.38), BMI (0.16), plantar arch temperature (0.11).
- **Random Forest:** Michigan Score (0.083), Total Tissue Thickness (HALLUX_R, 0.083), SESA_R (0.065).

- **Univariate ANOVA (Eta Squared):** Michigan Score (0.829), Total Tissue Thickness (HALLUX_R, 0.826) and TM5_L (0.809) had the largest effect sizes.

Feature importance rankings for clustering algorithms (with and without IWGDF grade) further confirmed the predictive role of Michigan Score, tissue thickness measures, and temperature-related parameters.

Including the IWGDF grade consistently improved clustering quality across both analyses (important parameters and all parameters), as evidenced by higher Silhouette and Calinski–Harabasz scores and lower Davies–Bouldin scores. The "important parameter" analysis yielded stronger separation (Silhouette: 0.55 vs. 0.22) and clearer cluster structures than the full-feature analysis, which may have been affected by noise from less-informative variables.

Feature importance analyses highlighted the Michigan Score, tissue thickness measures (HALLUX and TM5), and temperature metrics as key discriminators of IWGDF grades. Logistic regression with LASSO, Random Forest, and ANOVA consistently identified these variables as critical to group separation.

Overall, the results suggest that clustering using a focused set of clinically and biomechanically relevant parameters provides more interpretable and robust grouping than clustering with the entire feature set.

6.7 MANOVA with IWGDF Grade as Factor

To assess the overall effect of the IWGDF grade on the selected set of key biomechanical and clinical parameters — namely *MESI*, *Michigan Score*, *Pressure*, *Stiffness*, *Thickness*, *Amplitude*, and *Temperature* — a Multivariate Analysis of Variance (MANOVA) was performed.

The multivariate tests indicated a significant effect of IWGDF grade on the combined dependent variables. The results of the MANOVA are summarized below:

Table 7: Multivariate Test Statistics for MANOVA (Intercept)

Test	Value	Num DF	Den DF	F Value	Pr > F
Wilks' Lambda	0.0198	8	12	74.0930	<0.0001
Pillai's Trace	0.9802	8	12	74.0930	<0.0001
Hotelling–Lawley Trace	49.3953	8	12	74.0930	<0.0001
Roy's Greatest Root	49.3953	8	12	74.0930	<0.0001

Table 8: Multivariate Test Statistics for MANOVA (Effect of IWGDF Grade)

Test	Value	Num DF	Den DF	F Value	Pr > F
Wilks' Lambda	0.0560	8	12	25.2724	<0.0001
Pillai's Trace	0.9440	8	12	25.2724	<0.0001
Hotelling–Lawley Trace	16.8482	8	12	25.2724	<0.0001
Roy's Greatest Root	16.8482	8	12	25.2724	<0.0001

The significant multivariate effect ($p < 0.0001$) confirms that IWGDF grade strongly influences the combined biomechanical and clinical parameters, validating its use as a risk stratification tool but also highlighting the need for multidimensional assessment.

6.8 Correlation Analysis

Correlation analysis was performed to identify interdependencies among parameters, guiding the selection of variables for predictive models by highlighting those with strong relationships to DFU risk factors. Table 9 summarizes the strongest correlations observed between pairs of parameters across the dataset.

Parameter 1	Parameter 2	Correlation Coefficient
Grade IWGDF	Michigan Score (ok=13, risk=0)	-0.833
Grade IWGDF	Grade	1.000
Weight (kg)	BMI	0.867
MESI Ankle Pressure R	MESI Ankle Pressure L	0.842
MESI Ankle Pressure R	MESI Big Toe Systolic Pressure Index R	0.926
MESI Ankle Pressure L	MESI Big Toe Systolic Pressure Index L	0.896
Michigan Score (ok=13, risk=0)	Total Tissue Thickness HALLUX R	0.815
Michigan Score (ok=13, risk=0)	Total Tissue Thickness TM5 L	0.829
Michigan Score (ok=13, risk=0)	Grade	-0.833
Avg Pressure Max SESA R	Avg Pressure Max SESA L	0.872
Stiffness HALLUX R	Stiffness HALLUX L	0.897
US Thickness ED HALLUX L	US Thickness Hypodermis SESA L	-0.804
US Thickness Hypodermis SESA R	Total Tissue Thickness SESA R	0.991
US Thickness Hypodermis HALLUX R	Total Tissue Thickness HALLUX R	0.974
US Thickness Hypodermis TM5 R	Total Tissue Thickness TM5 R	0.970
US Thickness Hypodermis SESA L	US Thickness Hypodermis TM5 L	0.821
US Thickness Hypodermis SESA L	Total Tissue Thickness SESA L	0.993
US Thickness Hypodermis HALLUX L	Total Tissue Thickness HALLUX L	0.992
US Thickness Hypodermis TM5 L	Total Tissue Thickness SESA L	0.826
US Thickness Hypodermis TM5 L	Total Tissue Thickness TM5 L	0.988
Total Tissue Thickness SESA R	Total Tissue Thickness TM5 R	0.820
Total Tissue Thickness SESA R	Total Tissue Thickness TM5 L	0.834
Total Tissue Thickness HALLUX R	Total Tissue Thickness TM5 L	0.862
Total Tissue Thickness TM5 R	Total Tissue Thickness SESA L	0.922
Total Tissue Thickness TM5 R	Total Tissue Thickness TM5 L	0.975
Total Tissue Thickness TM5 R	ROC SESA L	0.813
Total Tissue Thickness SESA L	Total Tissue Thickness TM5 L	0.933
ROC SESA R	ROC SESA L	0.845
Temperature 5th Toe R	Temperature Forefoot R	0.927
Temperature 5th Toe R	Temperature Hallux L	0.881
Temperature 5th Toe R	Temperature Forefoot L	0.923
Temperature 5th Toe R	Average IR Temperature Foot R (Celsius)	0.855
Temperature 5th Toe R	Normalized Temperature R	-0.820
Temperature Plantar Arch R	Temperature Lateral Sole R	0.856
Temperature Plantar Arch R	Temperature Heel L	0.803
Temperature Plantar Arch R	Average IR Temperature Foot R (Celsius)	0.810
Temperature Lateral Sole R	Temperature Plantar Arch L	0.876
Temperature Lateral Sole R	Temperature Heel L	0.801

Parameter 1	Parameter 2	Correlation Coefficient
Temperature Lateral Sole R	Average IR Temperature Foot R (Celsius)	0.859
Temperature Lateral Sole R	Average IR Temperature Foot L (Celsius)	0.862
Temperature Forefoot R	Temperature Hallux L	0.907
Temperature Forefoot R	Temperature Plantar Arch L	0.845
Temperature Forefoot R	Temperature Forefoot L	0.933
Temperature Forefoot R	Average IR Temperature Foot R (Celsius)	0.947
Temperature Forefoot R	Average IR Temperature Foot L (Celsius)	0.885
Temperature Forefoot R	Temperature Difference Hand-Foot R	-0.811
Temperature Forefoot R	Temperature Difference Hand-Foot L	-0.819
Temperature Forefoot R	Normalized Temperature R	-0.852
Temperature Forefoot R	Normalized Temperature L	-0.843
Temperature Heel R	Temperature Plantar Arch L	0.805
Temperature Heel R	Average IR Temperature Foot R (Celsius)	0.827
Temperature Hallux L	Temperature Plantar Arch L	0.804
Temperature Hallux L	Temperature Forefoot L	0.955
Temperature Hallux L	Temperature Heel L	0.819
Temperature Hallux L	Average IR Temperature Foot R (Celsius)	0.879
Temperature Hallux L	Average IR Temperature Foot L (Celsius)	0.881
Temperature Hallux L	Temperature Difference Hand-Foot L	-0.803
Temperature Hallux L	Normalized Temperature R	-0.828
Temperature Hallux L	Normalized Temperature L	-0.823
Temperature Plantar Arch L	Temperature Forefoot L	0.823
Temperature Plantar Arch L	Temperature Heel L	0.894
Temperature Plantar Arch L	Average IR Temperature Foot R (Celsius)	0.832
Temperature Plantar Arch L	Average IR Temperature Foot L (Celsius)	0.833
Temperature Forefoot L	Temperature Heel L	0.811
Temperature Forefoot L	Average IR Temperature Foot R (Celsius)	0.881
Temperature Forefoot L	Average IR Temperature Foot L (Celsius)	0.894
Temperature Forefoot L	Temperature Difference Hand-Foot R	-0.806
Temperature Forefoot L	Temperature Difference Hand-Foot L	-0.828
Temperature Forefoot L	Normalized Temperature R	-0.840
Temperature Forefoot L	Normalized Temperature L	-0.845
Temperature Heel L	Average IR Temperature Foot L (Celsius)	0.816
Temperature Hand Mean D	Temperature Hand Mean L	0.896
Average IR Temperature Foot R (Celsius)	Average IR Temperature Foot L (Celsius)	0.947
Temperature Difference Hand-Foot R	Temperature Difference Hand-Foot L	0.950
Temperature Difference Hand-Foot R	Normalized Temperature R	0.996
Temperature Difference Hand-Foot R	Normalized Temperature L	0.936
Temperature Difference Hand-Foot L	Normalized Temperature R	0.954
Temperature Difference Hand-Foot L	Normalized Temperature L	0.996
Normalized Temperature R	Normalized Temperature L	0.947
SUDOSCAN Hand R	SUDOSCAN Hand L	0.976
SUDOSCAN Foot R	SUDOSCAN Foot L	0.994

Parameter 1	Parameter 2	Correlation Coefficient

Table 9: Correlation coefficients between key parameters.

6.8.0.1 Correlogram A correlogram is shown in Figure 12, displaying only strong correlations ($|r| > 0.7$) to reduce visual noise and emphasize the most relevant associations.

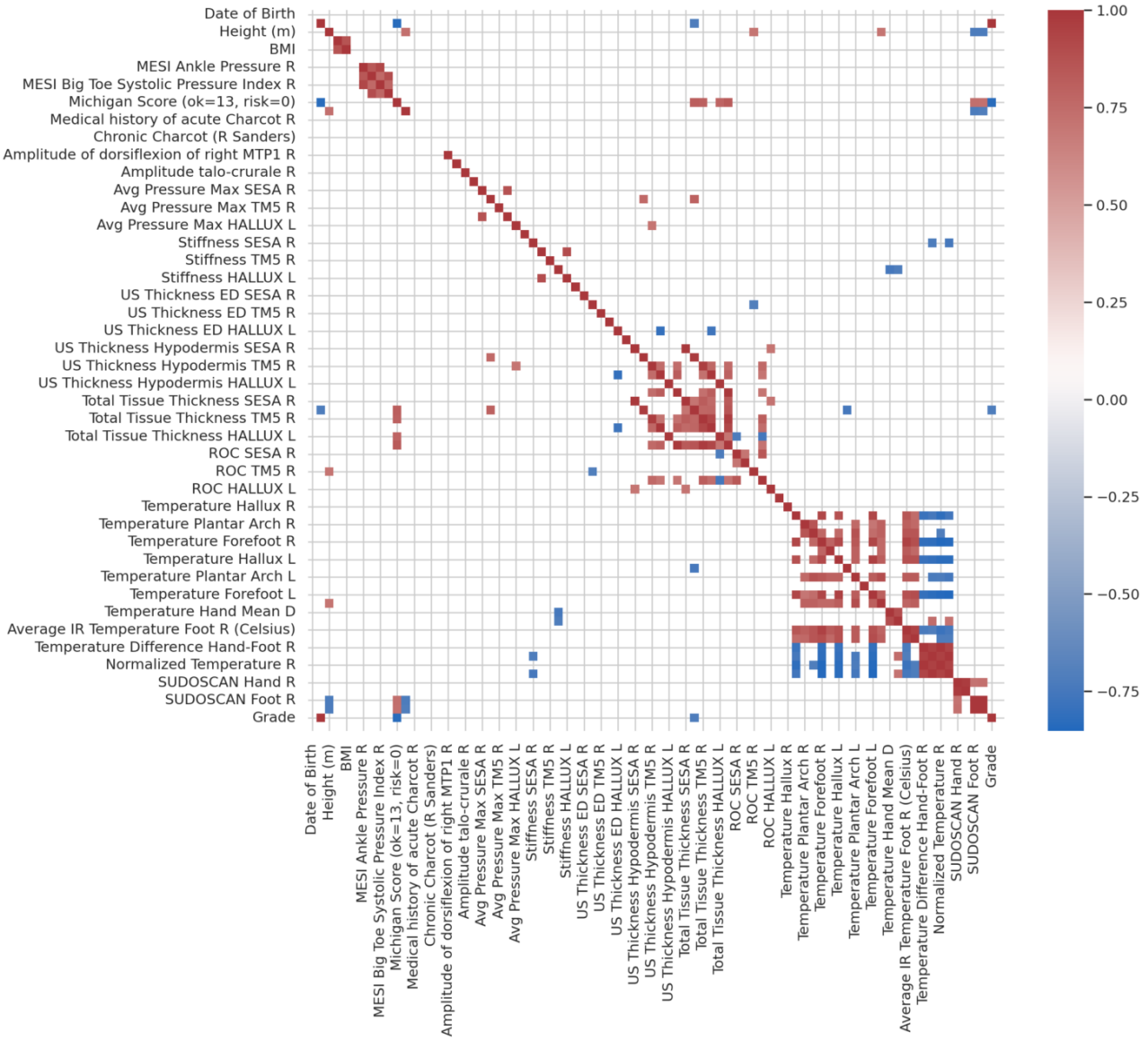


Figure 12: Correlogram showing correlations with $|r| > 0.7$ among selected DIAFOOT parameters.

Strong correlations between bilateral biomechanical parameters, tissue thickness ($\rho = 0.991$) and temperature ($\rho > 0.9$) demonstrated near-perfect positive correlations between feet in most patients.

A strong inverse correlation was observed between IWGDF risk grade and the Michigan Neuropathy Screening Instrument (MNSI) questionnaire score ($\rho = -0.833$). This finding reflects the opposing scales of these two measures:

- A **lower Michigan Score 1** (questionnaire) indicates **more severe neuropathy**, and therefore a **higher IWGDF grade**.
- A **higher Michigan Score 2** (clinical examination) also indicates **greater neuropathic impairment**, aligning with **higher IWGDF grades**.

These consistent yet directionally opposite relationships highlight the complementary nature of clinical scores and risk classification systems.

Taken together, the observed interdependencies suggest that integrating bilateral biomechanical measures with neuropathy screening scores in predictive models may enhance early detection of diabetic foot ulcer (DFU) risk and improve overall patient stratification.

7 Conclusion

The DIAFOOT dataset analysis provides critical insights into biomechanical and clinical factors associated with DFU risk, addressing the research question of how these factors can improve risk stratification and prediction. The logical progression of analyses—from descriptive statistics to clustering and correlations—reveals a multidimensional picture of DFU risk, with key findings driving actionable implications.

The cohort's characteristics (mean age 56.8 year, BMI 25.83 (kg/m^2), polarized IWGDF grades) highlight a diverse population with a high prevalence of overweight individuals and distinct risk profiles. Significant asymmetries in TM5 pressure ($p = 0.0038$), Hallux stiffness ($p = 0.0498$), Temperature plantar arch ($p < 0.006$) and Temperatures lateral sole ($p < 0.001$) suggest localized biomechanical and vascular changes that may precede ulceration, supporting the hypothesis that bilateral comparisons can enhance risk screening.

Clustering analyses demonstrated that a focused set of parameters (Michigan Score, tissue thickness, temperature) produced more robust and interpretable patient groups (Silhouette score = 0.553) than using all 87 variables (Silhouette score = 0.22). Including IWGDF grade improved clustering quality, but the moderate within-group inertia (43.875) and low Silhouette score for true grades (0.114) indicate that current risk classifications do not fully capture patient variability. Feature importance analyses (LASSO, Random Forest, ANOVA) consistently identified Michigan Score, tissue thickness (Hallux, TM5), and temperature as key predictors, suggesting their prioritization in future DFU risk models.

MANOVA confirmed the significant influence of IWGDF grade on multiple parameters ($p < 0.0001$), while correlation analysis revealed strong interdependencies (e.g., tissue thickness, $\rho = 0.991$; temperature, $\rho > 0.9$) and a notable inverse relationship between IWGDF grade and Michigan Score ($\rho = -0.833$). Comparisons with literature values validated the dataset's clinical relevance, with lower stiffness values potentially indicating early-stage neuropathy.

7.1 Key Implications

- **Enhanced Risk Stratification**: Michigan Score, tissue thickness, and temperature metrics are high-value markers for DFU risk, supporting their integration into predictive tools.
- **Asymmetry as a Risk Indicator**: Significant bilateral differences in pressure and temperature highlight the importance of side-specific analyses in clinical assessments.
- **Targeted Parameter Selection**: Clustering with key parameters outperforms comprehensive feature sets, reducing noise and improving model interpretability.
- **Limitations of Current Classifications**: The heterogeneity within IWGDF grades suggests a need for more nuanced risk stratification incorporating biomechanical data.

7.2 Limitations and Future Directions

The small sample size (n=21) and skewed grade distribution (few Grade 1–2 cases) limit statistical power and generalizability. Median imputation and data standardization may influence results for non-normal

parameters, and pooling bilateral data could mask critical asymmetries. Potential measurement biases (e.g., ultrasound or pressure variability) also warrant caution. Future studies should involve larger cohorts, standardized protocols, and longitudinal data to validate findings and incorporate dynamic measures like gait analysis. Advanced machine learning, such as deep learning, could further enhance feature extraction and prediction accuracy.

Despite these challenges, the integration of multiple biomechanical and clinical parameters appears promising for advancing personalized preventive strategies against diabetic foot ulcers.

7.3 Technical Reflections

The analytical workflow combined descriptive statistics, normality testing, group comparisons, clustering, and correlation analysis to comprehensively evaluate the DIAFOOT dataset. Implemented in Python, the pipeline leveraged libraries such as NumPy, Pandas, and Scikit-learn for efficient and reproducible processing of biomechanical and clinical data.

Clustering, effectively identified clinically meaningful patient subgroups.

MANOVA confirmed that IWGDF grade significantly influences multiple parameters, but the analysis was limited to a selected subset. Correlation analysis revealed strong interdependencies, although the focus on high thresholds ($|\rho| > 0.7$) may have excluded clinically relevant moderate associations.

7.4 Summary of Contributions

Regular collaboration via Google Meet (2–3 times weekly) ensured alignment with project goals. Mr. Haddad provided clinical expertise, validating parameter definitions and data accuracy. Mr. Rohan guided statistical method selection and analysis refinement, ensuring relevance to DFU risk assessment.

8 Reflection

The internship was a transformative experience, bridging theoretical data science with practical biomedical applications. Working with the DIAFOOT dataset honed my skills in Python-based analysis, statistical testing, and clustering, while overcoming challenges like missing data and time management. The project deepened my understanding of DFU risk factors and the value of interdisciplinary collaboration. Moving forward, I aim to leverage advanced machine learning and longitudinal data analysis to further contribute to preventive healthcare.

References

Armstrong, D. G., Boulton, A. J., and Bus, S. A. (2017). Diabetic foot ulcers and their recurrence. *New England Journal of Medicine*, 376(24):2367–2375.

Boulton, A. J., Vileikyte, L., Ragnarson-Tennvall, G., and Apelqvist, J. (2005). The global burden of diabetic foot disease. *The Lancet*, 366(9498):1719–1724.

Bus, S. A., Sacco, I. C. N., Monteiro-Soares, M., Raspovic, A., Paton, J., Rasmussen, A., Lavery, L. A., and van Netten, J. J. (2023). Guidelines on the prevention of foot ulcers in persons with diabetes (iwgdf 2023 update). *Diabetes/Metabolism Research and Reviews*, 40(3).

Bus, S. A., van Netten, J. J., Lavery, L. A., Monteiro-Soares, M., Rasmussen, A., Jubiz, Y., and Price, P. E. (2016). Iwgdf guidance on the prevention of foot ulcers in at-risk patients with diabetes. *Diabetes/Metabolism Research and Reviews*, 32(S1):16–24.

Cavanagh, P. R. and Bus, S. A. (2010). Off-loading the diabetic foot for ulcer prevention and healing. *Journal of Vascular Surgery*, 52(3):37S–43S.

Ceelen, K., Stekelenburg, A., Loerakker, S., Strijkers, G., Bader, D., Nicolay, K., Baaijens, F., and Oomens, C. (2008). Compression-induced damage and internal tissue strains are related. *Journal of Biomechanics*, 41(16):3399–3404.

Coleman, S., Gorecki, C., Nelson, E. A., Closs, S. J., Defloor, T., Halfens, R., Farrin, A., Brown, J., Schoonhoven, L., and Nixon, J. (2013). Patient risk factors for pressure ulcer development: Systematic review. *International Journal of Nursing Studies*, 50(7):974–1003.

Flynn, C. (2010). Finite element models of wound closure. *Journal of Tissue Viability*, 19(4):137–149.

Gefen, A., Brienza, D. M., Cuddigan, J., Haesler, E., and Kottner, J. (2021). Our contemporary understanding of the aetiology of pressure ulcers/pressure injuries. *International Wound Journal*, 19(3):692–704.

Hingorani, A., LaMuraglia, G. M., Henke, P., Meissner, M. H., Loretz, L., Zinszer, K. M., Driver, V. R., Frykberg, R., Carman, T. L., Marston, W., Mills, J. L., and Murad, M. H. (2016). The management of diabetic foot: A clinical practice guideline by the society for vascular surgery in collaboration with the american podiatric medical association and the society for vascular medicine. *Journal of Vascular Surgery*, 63(2):3S–21S.

International Diabetes Federation (2024). Number of adults (20–79 years) with diabetes in france (fr). IDF Diabetes Atlas, 10th edition. Accessed July 2025.

Jeffcoate, W. J. and Harding, K. G. (2003). Diabetic foot ulcers. *The Lancet*, 361(9368):1545–1551.

Keenan, B. E., Evans, S. L., and Oomens, C. W. (2022). A review of foot finite element modelling for pressure ulcer prevention in bedrest: Current perspectives and future recommendations. *Journal of Tissue Viability*, 31(1):73–83.

Li, Y., Wu, W., Xue, L., Zhao, T., Lu, Y., Qiao, X., and Ding, H. (2025). Plantar tissue characteristics in people with diabetes with and without peripheral neuropathy: A novel explanatory model for <scp>dpn</scp> risk assessment. *Journal of Diabetes*, 17(5).

Loerakker, S., Manders, E., Strijkers, G. J., Nicolay, K., Baaijens, F. P. T., Bader, D. L., and Oomens, C. W. J. (2010). The effects of deformation, ischemia, and reperfusion on the development of muscle damage during prolonged loading. *Journal of Applied Physiology*, 111(4):1168–1177.

Loerakker, S., Manders, E., Strijkers, G. J., Nicolay, K., Baaijens, F. P. T., Bader, D. L., and Oomens, C. W. J. (2011). The effects of deformation, ischemia, and reperfusion on the development of muscle damage during prolonged loading. *Journal of Applied Physiology*, 111(4):1168–1177.

Mens, M., Fassaert, T., Homan, J., Busch-Westbroek, T., Stufkens, S., Wellenberg, R., Streekstra, G., Bus, S., Nieuwdorp, M., and Maas, M. (2023). Sub-calcaneal plantar fat pad assessment using dual-energy computed tomography: First experience in the diabetic foot. *Clinical Biomechanics*, 110:106126.

Naemi, R., Chatzistergos, P., Sundar, L., Chockalingam, N., and Ramachandran, A. (2016). Differences in the mechanical characteristics of plantar soft tissue between ulcerated and non-ulcerated foot. *Journal of Diabetes and its Complications*, 30(7):1293–1299.

Naemi, R., Chatzistergos, P., Suresh, S., Sundar, L., Chockalingam, N., and Ramachandran, A. (2017). Can plantar soft tissue mechanics enhance prognosis of diabetic foot ulcer? *Diabetes Research and Clinical Practice*, 126:182–191.

O'Tuathail, C. and Taqi, R. (2011). Evaluation of three commonly used pressure ulcer risk assessment scales. *British Journal of Nursing*, 20(Sup2):S27–S34.

Paton, J., Bruce, G., Jones, R., and Stenhouse, E. (2011). Effectiveness of insoles used for the prevention of ulceration in the neuropathic diabetic foot: a systematic review. *Journal of Diabetes and its Complications*, 25(1):52–62.

Raspovic, A. and Landorf, K. B. (2014). A survey of offloading practices for diabetes-related plantar neuropathic foot ulcers. *Journal of Foot and Ankle Research*, 7(1).

World Health Organization (2016). Global report on diabetes. Technical report, World Health Organization. Published on the occasion of World Health Day 2016.