UNIVERSITI MALAYA

KIE4033

Data Analytics

Assignment 2

2025/2026

Developing AI-driven framework that automates both detection and clinical stratification of DFU classification and severity mapping

Name	*Matric Number*
Muhammad Rashidul Bukhari Bin Asan Mohamed	22004502/1

PREPARED FOR: ASSOCIATE PROF. IR. DR. CHOW CHEE ONN

Dataset Link (Kaggle): https://www.kaggle.com/datasets/laithjj/diabetic-foot-ulcer-dfu

Code Link (GitHub): https://github.com/Bukhari7703/KIE4033_Bukhari_Assignment2.git

Video Link (Microsoft Teams) [kindly use 1.2x speed]: KIE4033_Asgn2_22004502-20260118_042211-Meeting Recording.mp4

Abstract

Diabetic foot ulcers (DFUs) require precise monitoring to prevent amputations. Previous work by group RM-2 focused on basic geometric data such as perimeter, surface area, and dimensions while manually transmitting for clinical review. However, these measures are insufficient for early triage as they fail to reflect the biological state of the wound bed, such as infection or necrosis, which are powerful predictors of severity. This project proposes an AI-driven framework that automates both detection and clinical stratification. By engineering biomedical features like Colour Variance and Erythema Index, the model quantifies tissue complexity and inflammation. An optimized Random Forest classifier achieved 94% accuracy and a 0.99 AUC. This is enhanced by K-Means clustering, K=6, programmatically mapping wounds to the clinical Wagner Classification System. This enables automated triage and immediate risk alerts, reducing the need for constant clinician oversight.

Introduction

The management of Diabetic Foot Ulcers (DFUs) remains a critical challenge in modern medicine, where delayed intervention frequently results in irreversible tissue damage and lower-limb amputation. As a matter of fact, in areas where they are much needed, such as the state of Sabah, many DFU patients ignore the treatment due to distance and financial costs and come for specialist treatment only after the condition progresses to a life-threatening situation. The DFU parameters such as geometric data which are provided by the system designed by Group RM-2, are actually very superficial and do not measure the actual changes within the body which is the measure of necrotic changes and inflammation.

The primary objective is to automate the detection and clinical risk-stratification of ulcers using bio-medical feature engineering to provide immediate, Wagner-aligned triage for rural patients. This proposal brings forth an intelligent diagnostic system powered by AI that upgrades the RM-2 monitoring tool from a reactive wound tracker to a proactive medical companion by architecting innovative bio-medical feature development. Through its application of colour space evaluation using CIE LAB colour space analysis, it assesses Erythema and Necrosis amounts to provide a biological perspective on wound conditions. This enables informed decision-making through its binary classifier to distinguish normal skin from an ulceration site and an unsupervised learning classifier to define injury into six stages of severity. Against this analysis, it allows for referrals of highly predisposed clients to higher medical attention without necessarily having them under constant medical observation.

Problem Definition

Effective DFU management in rural healthcare environments is being hindered by the absence of a critical solution for automated triage or clinical decision support. Below are the three pillars that outline the problem being addressed using the diagnostic framework powered by Artificial Intelligence.

1. Limitations of Geometric-Only

The former monitoring device by Group RM-2 consists of calculating physical size information such as perimeter and surface area. Although these pieces of information are being transmitted and temporarily archived on a cloud for physician interpretation, from a medical viewpoint, these criteria are insufficient for stage one triage analysis because it cannot represent the biological status of the wound site. The presence of infection (erythema) and necrosis may be developing internally without manifesting a noticeable effect on wound size.

2. The Decision-Making Triage Bottleneck

Device by RM-2 functions as a tool for the collection of data for synchronization with the cloud for the doctor's review. It causes a bottleneck situation for the decision-making process in an area where the number of patients can be high, and the availability of medical practitioners can be low. The device does not have the automatic grading system for risks; hence the need for the doctor, on a physical worksheet, to quantify the risks associated with each.

3. Environmental and User Variability in AI for Medicine

A challenge associated with remote monitoring is the high variability of the light, angles, and skin tones, which RM-2 pointed out as a significant issue related to "Consistency across users." A strong solution must shift from the reliance on simple pixel measurement to invariant bio-medical feature development. This would ensure the precision of the clinical result regardless of the lighting conditions within the rural home environment.

To overcome these challenges, this research has set the following objectives related to the clinical scenario of rural DFU patients:

- **Automate Detection using Machine Learning**: Develop an optimized Random Forest model that can detect ulcers with high sensitivity (0.94 accuracy).
- **Automated Risk Strategy**: Implement K-Means clustering to automatically stratify the severity of ulcers into six levels.
- **Direct Decision Support**: Directly map the severity levels to the Wagner Classification System, allowing the immediate prioritizing of critical case.

Data Description & Exploratory Data Analysis

Data Description and Preprocessing

The foundation of this diagnostic framework relies on a medically validated dataset from Kaggle, Diabetic Foot Ulcer DFU. The project uses the **Patches (Ulcer + Healthy)** directory of 512 and 543 pictures respectively, which includes high-resolution localized pictures of skin tissue, and the **TestSet** directory, containing ulcer pictures for external testing of the model of 167 pictures.

First, the pictures data are pre-processed according to two different pipelines to make sure that the analysis will be rigorous and well-validated. The first one is **Internal Dataset** (0_dfu_features.csv). It compromises both Patches folder where the **Ulcer** skin is mapped as 1 and **Healthy** skin is mapped as 0 under column ulcer present. This set is used for the primary training and internal testing of the classification model and to perform risk stratification. The second one is **External Validation Set** (1_dfu_features_testset.csv): A separate collection of ulcer images only mapped to 1 used exclusively to validate the robustness of the classification model.

To address the Variability in Wound Appearance and Inconsistency problems identified by RM-2, the following multi-level preprocessing process was developed:

- **Image Standardization and Denoising**: All images will be uniformed to a standard size of 256 x 256 pixels. A Median Filter is applied because it is excellent at removing random bright or dark pixels that appear as speckles and the wound edge will be preserved, which is a key in distinguishing the boundaries
- **Clinical Colour Space Transformation**: Contrary to conventional RGB colour space, pictures are transformed into CIE LAB colour space. This ensures that Lightness, which identifies necrosis, as well as Red Green axis to quantify erythema. This makes the system less affected by varying light conditions prevalent in rural homes.
- **Morphological Segmentation:** Otsu's Thresholding on the Lightness channel is employed for automatically extracting the major wound region from the surrounding healthy skin.
- **Dealing with Missing Values in Outlier Data**: Effective error handling is taken care of in this process, meaning that if there exists any image where it is not possible to segment out the main affected area, then it is avoided completely with the following code: (in case of not props: return None).

The transformation from raw images to csv dataset is achieved through the extraction of seven distinct features which is shown in Table 1.

Group	Features	Purpose
Geometric Features	Area, Perimeter, Solidity & Eccentricity	Calculated to refine the basic metrics used by Group RM-2
Bio-Medical Features	Erythema Index (mean A-channel) Necrosis Index (mean L-channel)	Calculated to provide the biological context

Tissue Complexity	Colour Variance (standard deviation of the image)	Used as digital fingerprint for the heterogeneity of complex ulcerous tissue
Target Feature	Presence of Ulcer	Binary value indicating ulcer present: 1 and ulcer absent: 0

Table 1: Feature Extraction

Exploratory Data Analysis

The EDA phase was conducted to identify the statistical patterns that distinguish healthy skin from ulcerous tissue. The univariate distributions reveal the spread and density of the seven extracted features across the total of 1,055 data points in Figure 1.

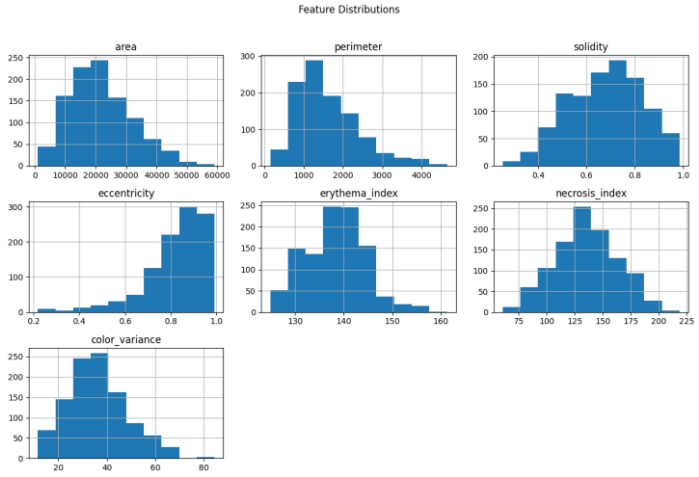

Figure 1: Univariate Feature Distribution

Metrics like area and perimeter show a wide range of values with a right-skewed distribution, confirming the high variability in ulcer sizes that makes simple geometric tracking unreliable. Features such as erythema_index and necrosis_index exhibit more Gaussian-like distributions, which are generally more stable for machine learning models. color_variance shows a distinct spread, providing a heterogeneous nature of ulcerated tissue.

A comparative analysis using side-by-side boxplots in Figure 2 highlights the effectiveness of the engineered features in separating the two classes. The colour variance boxplot shows the most significant separation. Ulcers exhibit a significantly higher mean colour variance compared to healthy skin, confirming that the chaotic colours of wound beds are a primary diagnostic driver. Ulcerous regions show a lower necrosis index and lower solidity compared to healthy skin. Clinically, this reflects the darker tissue tones and the more irregular, non-convex boundaries typical of active diabetic ulcers. Area shows a high degree of overlap between the two classes statistically proving why the junior team's reliance on simple size was insufficient for automated detection.

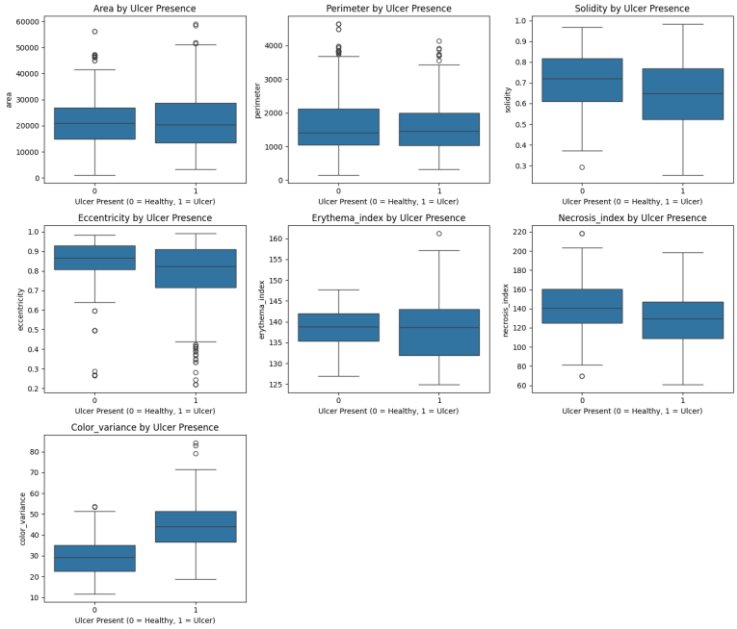

Figure 2: Boxplot Ulcer & Healthy

The correlation heatmap in Figure 3 was utilized to identify relationships between features and ensure the model is not trained on redundant data. Area vs Perimeter show a moderate positive correlation of 0.48, as larger wounds naturally tend to have longer boundaries. However, the correlation is not high enough to warrant dropping either, as they capture different aspects of wound shape. Erythema vs Colour Variance show the strongest positive correlation of 0.39 among the biomedical features. It suggests that as inflammation (redness) increases, the wound bed often becomes more heterogeneous and complex. Eccentricity vs. Area shows a moderate negative correlation of -0.50 suggests that as ulcers grow larger, they tend to become less elongated and more circular or irregular in shape. For Area vs. Colour Variance, the corelation is near-zero of 0.038 between the size of a wound and its tissue complexity. This statistically proves that a small ulcer can be biologically more severe than a large, stable one.

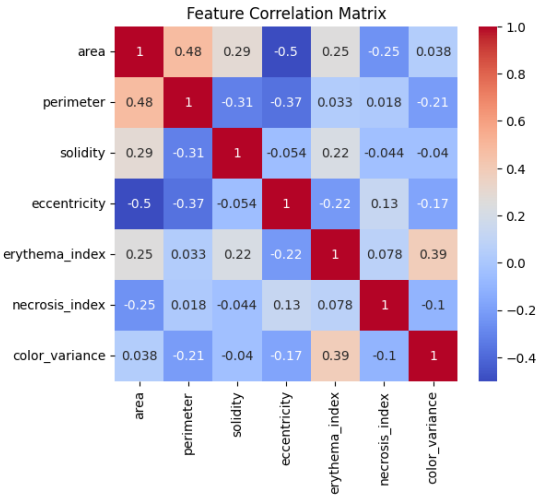

Figure 3: Correlation Heatmap

Proposed AI-based Solution Concept

The proposed solution addresses the limitations of the RM-2 monitoring system by introducing a sophisticated, two-tiered analytical framework that combines **Supervised Learning** for automated ulcer detection and **Unsupervised Learning** for severity-based triage and Wagner mapping.

1. Automated Ulcer Detection (Supervised Learning)

The core of the detection system is an Optimized Random Forest (RF) classifier. Hyperparameter optimization was conducted to move beyond basic modelling, GridSearchCV was utilized and the system identified n_estimators = 50 as part of the best parameters, proving that a lean, efficient forest is sufficient for high-accuracy detection on mobile health devices.

The RF model was selected as the core engine due to its Bootstrap Aggregation technique, which minimizes the risk of overfitting by training multiple decision trees on different data subsets. It also works well with heterogeneous datasets and reduces overfitting. The model also ignores raw pixel data in favour of the extracted features. This rationale ensures the AI prioritizes Clinical Data such as tissue complexity and redness over superficial environmental factors like lighting or background noise.

The proposed solution transitions from the passive tracking of physical dimensions to a proactive, biologically aware diagnostic framework. The model is trained with all features except the target feature as stated in Table 1. This model identifies ulcers by prioritizing colour variance and the erythema index that enables the system to detect early-stage ulcers that might be small but high in biological complexity which RM-2 ignored.

2. Severity based Triage and Wagner Mapping (Unsupervised Learning)

The second stage of the AI framework addresses the clinical need for automated risk stratification, moving beyond detection to provide a Severity Risk label for confirmed ulcers. This solution is designed to act as a digital triage assistant.

Unlike classification, which requires pre-labelled data, we implemented K-Means Clustering to discover the natural biological groupings within the ulcerous tissue data. This approach is innovative because it allows the AI to identify subtle variations in wound pathology such as the transition from superficial inflammation to deep necrosis that might be missed by manual inspection of geometric metrics.

To achieve high medical accuracy, the clustering engine utilizes the two most significant diagnostic features which is Colour Variance and the Erythema Index as shown in Figure 6. Because these features have different numerical scales, we applied StandardScaler to normalize the data. This step is critical in AI for Medicine to ensure that one metric does not mathematically overwhelm a more clinically significant one.

To ensure the triage levels are statistically sound and not arbitrary, the solution employs a dual-validation methodology of Elbow Method (WCSS) and Silhouette Analysis. Elbow Method is used to identify the inertia point of the datasets where adding more clusters no longer significantly improves the model's understanding of the data while Silhouette Analysis measures how well-separated and distinct the risk groups are. The system programmatically identified K=6 as the optimal split. This provides a direct, data-driven bridge to the Wagner Classification System, a global standard for DFU severity grading.

RM-2 product suffered from Surface-Level Monitoring**,** failing to measure internal biological changes. Presenting simple physical parameters provides no prioritization for high-risk patients. By calculating the mean values of each cluster, the AI identifies when a wound transitions from manageable to critical.

As for the mapping of Wagner scale to each cluster, colour variance is chosen to be mapped linearly with Wagner scale. The Wagner Scale is traditionally defined by the depth of tissue involvement. In a remote monitoring context where physical probing is impossible, colour indices (colour variance) act as non-invasive proxies for these biological states. The model also can ignore the distraction of size and identifies the tissue quality. This prevents a dangerous situation where a small but necrotic wound is dismissed simply because its area hasn't increased. Moreover, Wagner Grades 4 and 5 are defined gangrene. This is mathematically identified by extreme colour variance, as necrotic tissue presents as a heterogeneous mix of black, yellow and dark red.

The Wagner Scale with its explanation is stated in Table 2.

Wagner Grade	Diagnosis
0	Pre-ulcerative / Intact Skin
1	Superficial Ulcer
2	Deep Ulcer
3	Deep Abscess / Osteomyelitis
4	Localized Gangrene
5	Extensive Gangrene

Table 2: Wagner Scale

Results & Discussion

1. Automated Ulcer Detection (Supervised Learning)

RF was trained and tested on an 80:20 split. The model has a level of accuracy at 0.94, which is complemented by high Recall and Precision at 0.93 and 0.95 respectively for ulcerous tissue. This confirms that the system is very efficient at suppressing False Negatives which is crucial for patient safety in rural areas, and at the same time, maintaining a high level of diagnostic confidence. Crucially, the model identified 102 ulcers with only 8 missed cases, ensuring high safety for rural patients who need urgent care as shown in Figure 4. The model is also externally validated using testset dataset and achieved a 0.93 accuracy. Furthermore, the model maintained high accuracy of 0.93 during External Validation, proving it can generalize to images from different users and camera qualities from different users.

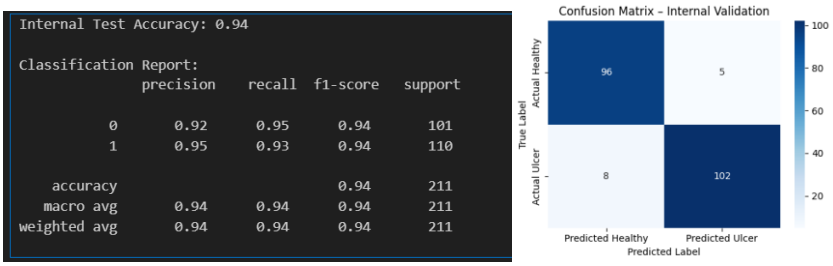

Figure 4: Classification Report & Confusion Matrix

Receiver Operating Characteristic, (ROC) curve in Figure 5 was plotted to measure the trade-off between sensitivity and false alarms. The AUC of 0.9887 indicates nearly perfect discriminative power. This model is robust enough to separate healthy skin from ulcers across a wide range of patient skin tones and wound types. This result also confirms that the features engineered in the CIE LAB space provide the model with enough biological information to separate classes with extreme confidence.

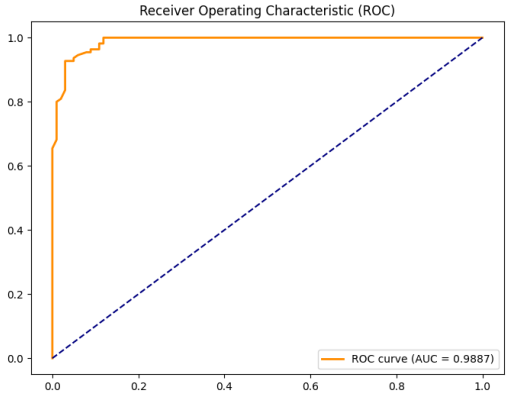

Figure 5: ROC Curve

Feature importance was plotted in Figure 6 for all seven features to see exactly which features drive the RF model. Colour Variance emerged as the most critical feature, with a Gini Importance score exceeding 0.40. Ulcer tissue is biologically heterogeneous, consisting of a mix of red granulation, yellow slough, and black eschar. The Erythema Index (Redness/Inflammation) is the second most influential feature. This justifies the project rational where the model ignores geometrical data and focuses on the biological fingerprints of the wound such as Colour Variance and Erythema Index, which are the true indicators of ulcer presence. This validates our rationale that the tissue complexity is significantly more powerful predictor of ulcer presence than geometric size. This shift enables the monitoring device to identify early-stage ulcers that may be small in area but high in tissue complexity.

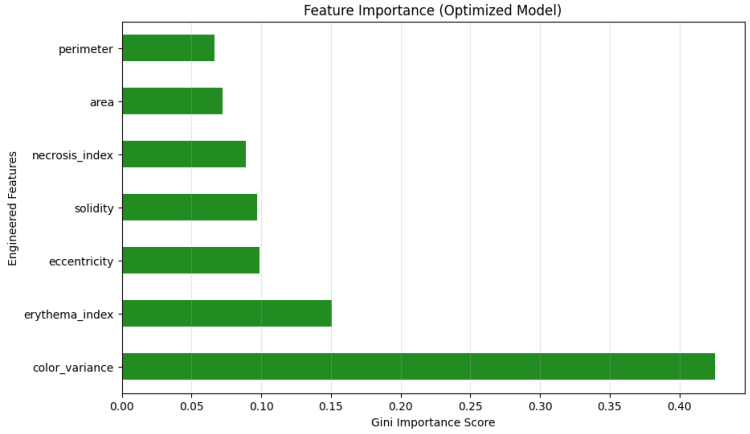

Figure 6: Feature Importance

2. Severity based Triage and Wagner Mapping (Unsupervised Learning)

The unsupervised learning phase of this project transforms raw bio-medical data into a structured clinical triage system. Only 2 features which is Colour Variance and the Erythema Index is used to perform K-Means clustering. By grouping confirmed ulcers based on biological fingerprints rather than just physical size.

Figure 7 shows Elbow (WCSS) and Silhouette Score plot to mathematical validate the number of clusters, best_K for clustering. This best_K will be used simultaneously as the risk categories. The WCSS curve shows an elbow at K = 6, indicating the point where adding further clusters provides diminishing returns in explaining data variance. The peak silhouette score of approximately 0.395 occurs at K = 6, confirming that these six groupings provide the highest degree of cluster separation and internal cohesion. This programmatically deduced best_K = 6, matching the Wagner Classification System (Grade 0–5) very well. The ground for this in the AI's logic thus largely rests on established standards in medicine.

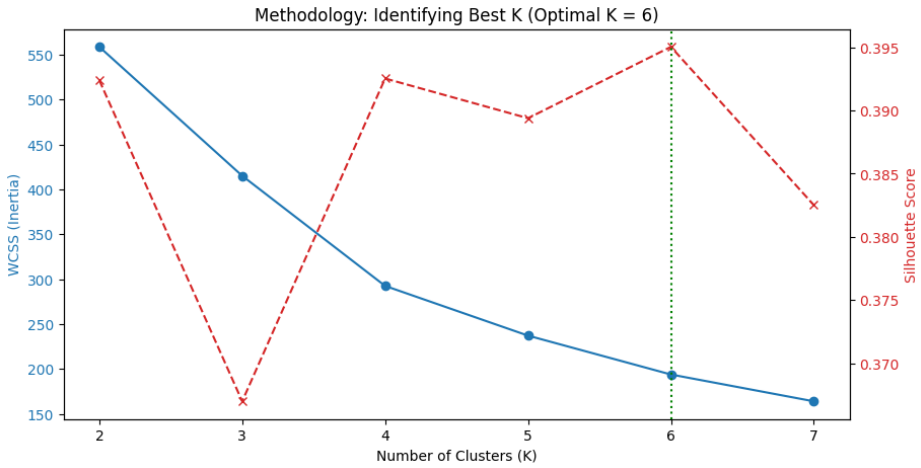

Figure 7: WCSS and Silhouette Score plot

Figure 8 shows scatter plot of the 2 chosen features. The plot clearly visualizes how wounds migrate from the bottom-left which is low risk to the top-right which is extreme risk as tissue complexity and inflammation increase. The plot also shows the cluster is separated into 6 different clusters with no visible overlapping. This result is good as the model can identify the groups distinctively with high accuracy due to no visible overlapping.

While the previous RM-2 system provided Passive Monitoring (geometric size), this AI-driven stratification offers proactive triage of severity clusters directly mapping to Wagner grading system. The system can instantly flag high-risk uploads ensuring that rural patients in Grade 4 or 5 clusters receive life-saving referrals within minutes of image upload.

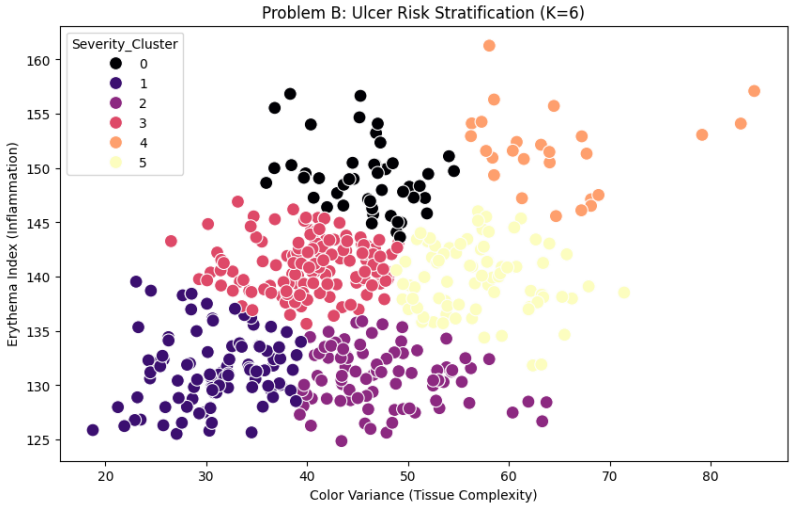

Figure 8: Scatter plot of Severity Cluster

The analysis of cluster centroids in Figure 9 reveals a clear linear progression of colour variance severity, allowing for a professional mapping to the Wagner Scale. Clusters 4 maps to Grade 5 because it exhibits extreme Colour Variance, signalling severe tissue heterogeneity

and necrosis typical of gangrene. On the other hand, Cluster 1 represents the lowest Wagner Grade with a low colour variance of 30.86, indicating no widespread infection. Cluster 0 maps to Grade 2 due moderate colour variance but the second-highest Erythema Index. This directly corelates to Deep Ulcers where inflammation is the primary risk factor, requiring urgent antibiotic intervention before necrosis expands.

# Severity_Cluster	# color_variance	# erythema_index	# Wagner_Grade	Clinical_Diagnosis
1	30.88533459628948	131.31511983785543	0	Grade 0: Pre-ulcerative / Intact Skin
3	40.62480468652054	141.06661433389743	1	Grade 1: Superficial Ulcer
0	45.793149015683014	149.17744266285615	2	Grade 2: Deep Ulcer (Tendon/Capsule)
2	47.47838073339367	130.69061296064774	3	Grade 3: Deep Abscess / Osteomyelitis
5	56.90425631837382	139.86247565947383	4	Grade 4: Localized Gangrene
4	64.08303719328129	151.52416330973307	5	Grade 5: Extensive Gangrene

Figure 9: Wagner Mapping to Cluster

Conclusion

The successful development of this diagnostic framework based on the use of the AI technology closes the gap of making the Group RM-2 monitoring tool a proactive health assistant and a proactive health partner by concentrating on bio-medical feature engineering as opposed to the common focus on the superficial physical aspects.

Summary of Results

Both classification and clustering using Random Forest and K-means demonstrated exceptional data driven solutions, diagnostics and stratifying capabilities.

The Optimized Random Forest model for automated detection reached an internal accuracy of 94% with an AUC near perfection of 0.9887. Furthermore, a very high recall of 0.93 makes sure that 93% of the active ulcers are correctly caught and decreases the risk of missing a diagnosis in unsupervised home settings significantly.

The K-means algorithm for severity and Wagner mapping uses K = 6. The system mathematically justifies the existence of six distinct risk profiles which honours the Wagner Classification System by identifying a clear linear progression from Grade 0 with the lowest tissue complexity to Grade 5 with maximum Colour Variance of 64.08 and extreme Erythema of 151.52 mapped to Cluster 4. The model successfully flagged Cluster 0 (Grade 2) as a high-inflammation risk zone, providing an early warning for deep-seated infection that traditional geometric tracking would have overlooked.

Moreover, Colour Variance is proven a vastly more powerful diagnostic driver than any other feature calculated due to its large Gini important score in determining the presence of ulcer. This enables the system to detect silent necrosis that geometric tools would ignore.

Healthcare & Business Intelligence

The integration of this AI framework into the RM-2 monitoring device transitions the product into a high-value clinical intelligence platform. The following actionable insights demonstrate the project's ability to drive both healthcare efficacy and business-level strategic outcomes.

First is reduction of cost of care. The AI model serves as a financial safeguard for patients as well as the healthcare system. By programmatically identifying patients the severity grades, the model serves to prevent low risk patients with limited means from having to pay the price of high travel costs and lost wages for visiting city-based specialists. This will save patient money and reduces the operational burden on hospital facilities.

Second is specialist resource optimisation. Specialist physicians can move from a traditional first-come, first-served system to a very efficient review system. Specialists should primarily attend to those patients with high levels of disease severity, (Grades 4 and 5), and in

doing so, they should be able to extend the level of patients they remotely monitor. Furthermore, aggregated cluster data allows for the strategic deployment of specialized podiatric teams to rural areas identified as having the highest density of amputation risks.

Third, inventory management is transformed from reactive to predictive. By quantifying the exact number of patients at each severity level, health departments can optimize the procurement of high-cost advanced wound care supplies, such as silver dressings or vacuum-assisted closure kits. This ensures that life-saving materials are restocked based on real-time biological data before critical stock-outs occur.

The final one is market differentiation and strategic growth. Unlike traditional solutions that concentrate on common physical parameters, the offered solution enables Wagner-attuned intelligence. This alignment of healthcare standards renders the device a decision-ready solution to healthcare providers around the world, paving the way for a large-scale government contract to enter untapped healthcare markets.

References

Laith Alzubaidi, M. A. (2020). DFU_QUTNet: diabetic foot ulcer classification using novel deep convolutional neural network. *Multimedia Tools and Applications*, 79, 15655 – 15677.

Laith Alzubaidi, M. A. (2020). Towards a Better Understanding of Transfer Learning for Medical Imaging: A Case Study. *Applied Science*, 10(13), 4523.

Laith Alzubaidi, M. A. (2022). Robust application of new deep learning tools: an experimental study in medical imaging. *Multimedia Tools and Applications*, 81, 13289 – 13317.

Priti Shah, R. I. (2022). Wagner's Classification as a Tool for Treating Diabetic Foot Ulcers: Our Observations at a Suburban Teaching Hospital. *Cureus*, 14(1).