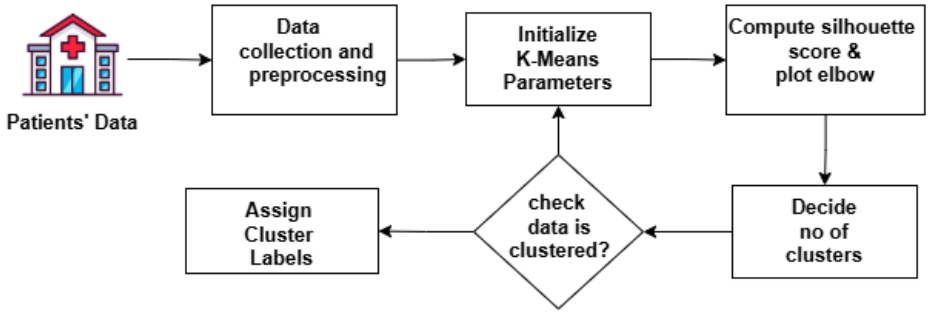
1 Supplementary material of Literature Review

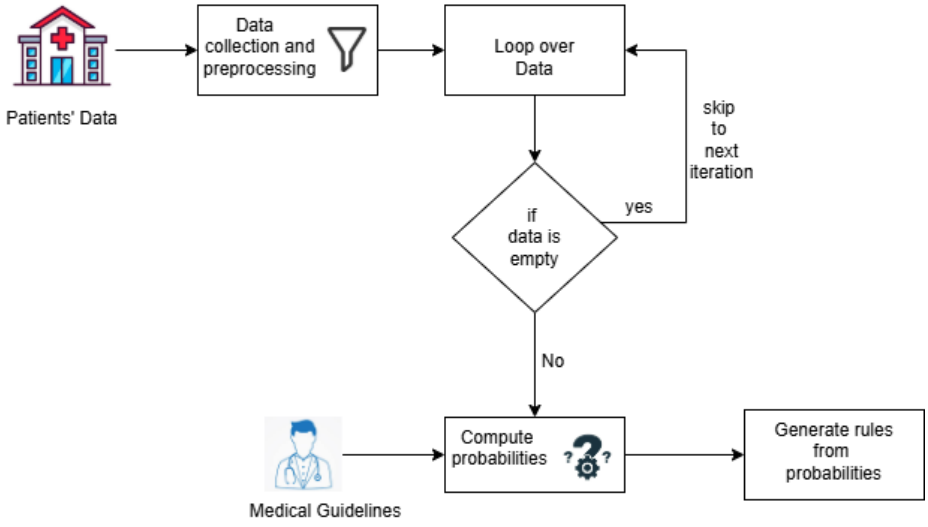
Supplementary Table 1: AI-Based Techniques for Burn Diagnosis and Analysis

Citation	Technique	Limitations	Accuracy (%)
[1]	ResNeXt, AlexNet	The model prioritizes local feature extraction and does not adequately consider global contextual information. It struggles to differentiate between deep dermal and full-thickness burns due to similar appearance, limiting classification accuracy.	84.31, 70.57
[2]	Random Forest, SVM	Trained on data from multiple ICUs but does not perform equally across units. Variability in clinical notes challenges robustness and generalizability.	84, 75
[3]	Unweighted/Weighted Mean Accuracy	Most models trained on historical datasets and not clinically validated. ML algorithms struggle to replicate nuanced decisions of experienced physicians.	81.54, 83.84
[4]	CNN, CNN-BAM	Trained on custom datasets, making models less applicable in other clinical environments. Traditional tools like LDI sometimes outperform AI-based assessments.	85, 91.6
[5]	SVM	Small dataset leads to overfitting, limiting real-world reliability.	88
[6]	VGG16, BNeXt	High computational resource requirements limit deployment to well-equipped hospitals.	78.23, 99.67
[7]	CNN	Does not support TBSA calculation; computationally intensive, limiting rural or real-time use.	93.3
[8]	FCNN	Requires powerful hardware, limiting real-time clinical usability. Lack of lightweight implementations restricts deployment outside major hospitals.	79.4
[9]	YOLO V7	Struggles to generalize across diverse skin colors, textures, and lighting conditions.	75.12
[10]	ResNet50, DL4 Burn	Trained only on a single dataset; not evaluated on independent datasets, raising concerns about generalizability.	81, 93.8
[11]	DeeplabV3+	Restricted to analyzing burn images captured within first 48 hours; performance may degrade afterward.	98
[12]	Random Forest, KNN	Traditional ML models less accurate for complex image classifications; deep learning requires powerful hardware, less practical for low-resource environments.	71.3, 66.67
[13]	Mask R-CNN, Dense Pose Model	High accuracy but not optimized for real-time scenarios; performs poorly for images from different angles.	87.24, 85.26
[14]	Random Forest	Small dataset increases overfitting; may fail on new patient cases.	92.2
[15]	CNN, ResNet50, GANs	Uses only 2D images; shows higher training than testing accuracy, indicating overfitting and limited generalizability.	96.88, 93.6

2 Supplimental matarial and Methods

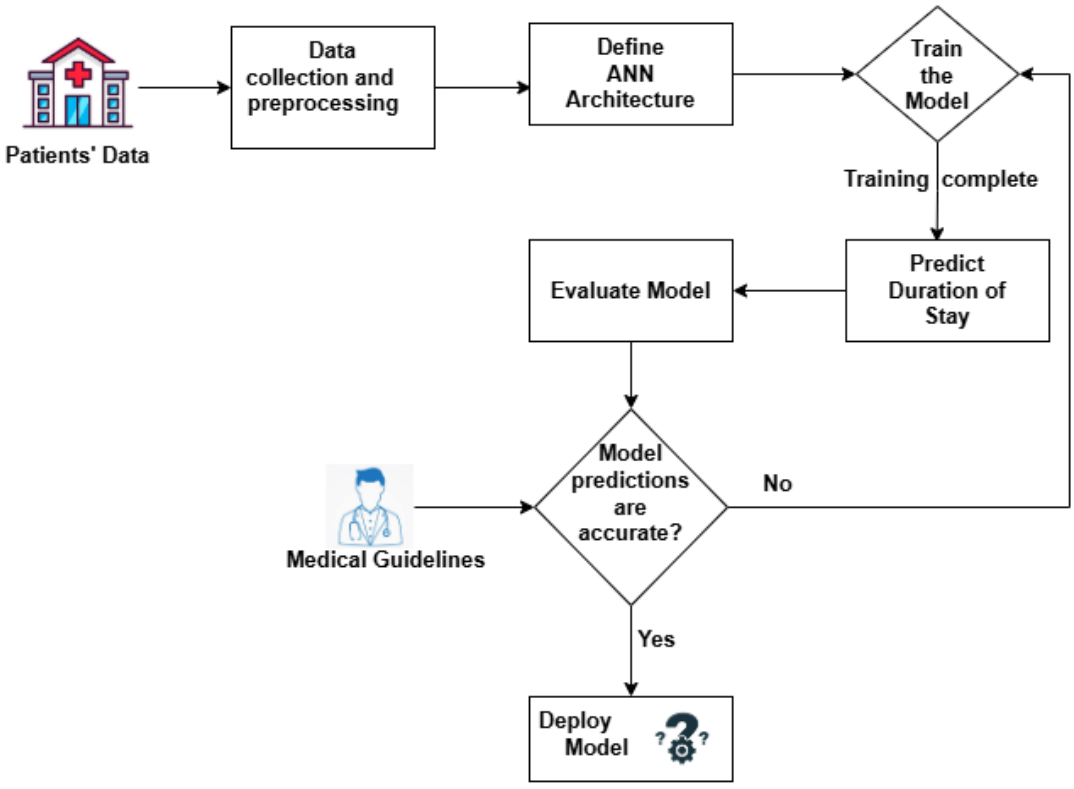


Supplementary Figure 1: Methodology of K-Means Clustering



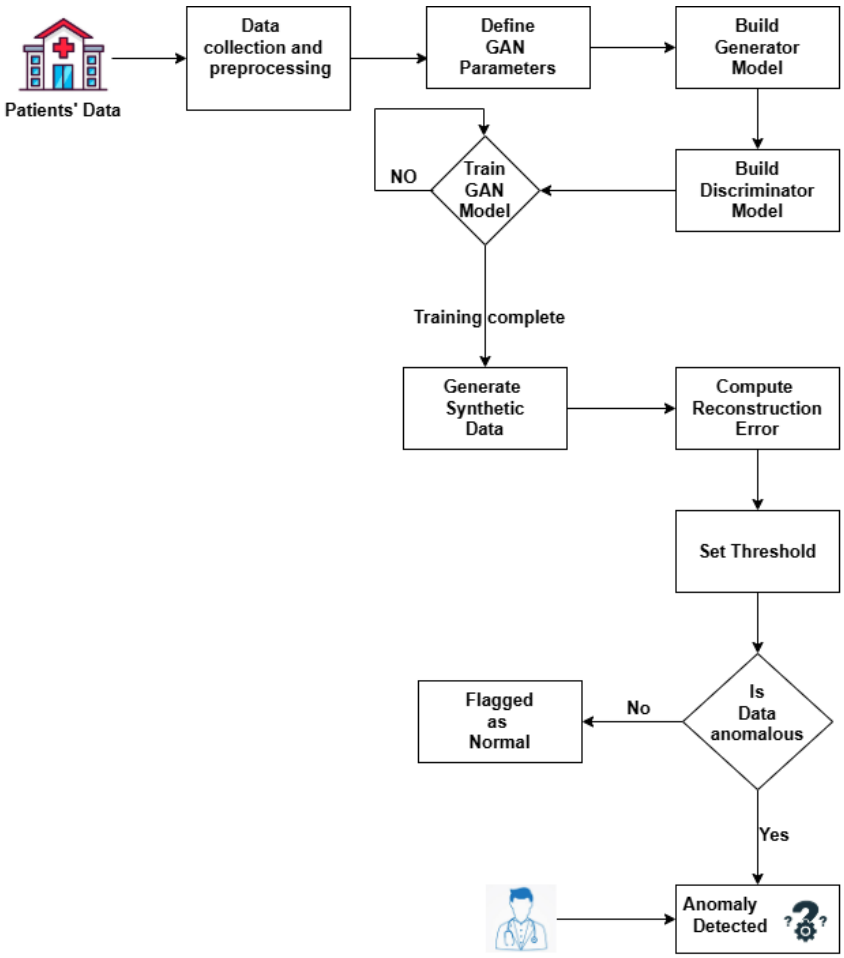
Supplementary Figure 2: Rules Generation.

Figure 2 shows methodology of Bayesian Theorem used for outcome prediction.



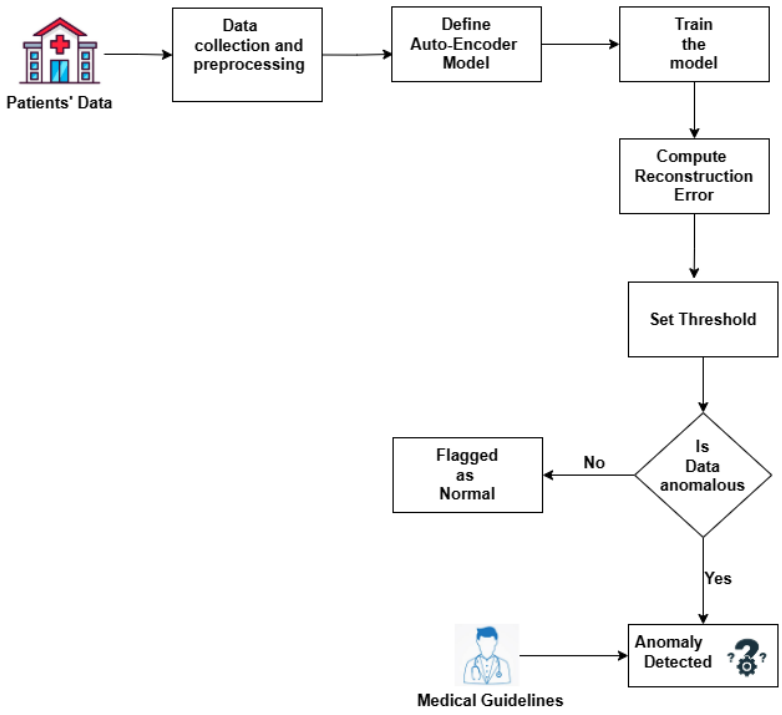
Supplementary Figure 3: ANN Methodology

4 Figure 3 shows methodology of Artificial neural network used for Inpatient stay prediction.



Supplementary Figure 4: Anomaly Detection Using GANs

Figure 4 shows methodology of GAN used for anomaly detection.



Supplementary Figure 5: Anomaly Detection Using Auto-Encoders

Figure 5 shows methodology of Auto-Encoder used for anomaly detection.,

---

Table 4 lists all notations used in Algorithm 1 , which outlines the complete K-Means clustering workflow, including data preprocessing, evaluating multiple cluster counts, determining the optimal number of clusters, and assigning cluster labels to the dataset.

The notations used in the ANN-based hospital stay prediction model are summarized in Table 3. The dataset D is split into input features X and target variable y, and the model predicts the patient's length of stay P. Algorithm 2 outlines the complete workflow, including data preprocessing, feature scaling, ANN architecture with three hidden layers, model compilation, training, evaluation, and prediction. Together, the table and algorithm provide a clear overview of the input, output, and procedural steps employed in the ANN framework.

Algorithm 3 presents the workflow of the Autoencoder-Based Anomaly Detection algorithm alongside the notations used. That outlines the steps from data preprocessing, autoencoder model construction, training, and computation of reconstruction errors, to the final identification of anomalies. The corresponding notation table (Figure 2) provides clear definitions of all variables and parameters used in the algorithm, ensuring easy reference and understanding.

Algorithm 3: Autoencoder-Based Anomaly Detection

Input: Dataset D, Learning rate η, Batch size B, Maximum epochs E, Patience P
Output: Detected anomalies A

1. Load the dataset D
2. Preprocess the dataset
3. Define Autoencoder architecture:
   - Input layer: x
   - Encoding layers using f_θ
   - Bottleneck layer
   - Decoding layers using g_θ
   - Output layer: x^*
4. Compile model and initialize EarlyStopping with patience P
5. Train autoencoder for E epochs with batch size B
6. Compute reconstruction error E_r
7. Define threshold T and identify anomalies A

Supplementary Table 2: Notations used in Autoencoder-Based Anomaly Detection

Notation	Description
D	Input dataset
η	Learning rate
B	Batch size
E	Maximum number of epochs
P	Patience
x	Input feature vector
f_θ	Encoder function
g_θ	Decoder function
x^*	Reconstructed output
E_r	Reconstruction error
T	Threshold
A	Detected anomalies

Algorithm 2: ANN-Based Hospital Stay Prediction

Input: Patient dataset D
Output: Predicted duration P

1. Load and preprocess dataset D
2. Split into features X and target y
3. Partition data into training, validation, and test sets
4. Scale features using MinMaxScaler
5. Build ANN: Input $\rightarrow$ Hidden layers (128, 64, 32, ReLU) $\rightarrow$ Output (linear)
6. Compile model with Adam optimizer and Huber loss
7. Train model for 100 epochs with validation
8. Evaluate model performance on test set
9. Predict duration: $P \leftarrow \text{Model.predict}(X_{\text{test}})$
10. **Return:** P

Supplementary Table 3: Notation Table for ANN-Based Length of Stay Prediction

Notation	Description
D	Dataset
X	Input features
y	Target variable (length of stay)
P	Predicted duration of stay

2.1 Evaluation Metrics

The two model variants, **Initial Model** and **Fine-tuned Model**, are evaluated using standard multi-class classification metrics: *Accuracy*, *Precision*, *Recall*, *F1-score*, and *Validation Loss*. Since this is a **multi-class problem** (1st, 2nd, and 3rd degree burns), metrics are computed for each class and summarized using **macro** and **weighted averages**.

1. Accuracy

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} = \frac{\sum_{i=1}^{C} TP_i}{\sum_{i=1}^{C} (TP_i + FP_i + FN_i)}$$

where C is the number of classes, TP_i is true positives for class i, FP_i is false positives, and FN_i is false negatives.

2. Precision

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i}$$

Algorithm 1: K-Means Clustering

Input: D (dataset), K (range of clusters), r (random seed)
Output: C (final cluster assignments)

1. Load and preprocess dataset
2. Initialize parameters: D, K, and r
3. **for each** $k \in K$ **do**
4. Fit K-Means model with k clusters
5. Compute silhouette score S_k
6. Store S_k
7. **end for**
8. Plot silhouette score vs. number of clusters
9. Identify optimal k with highest silhouette score
10. **for each** $k \in \{1, 2, \ldots, 10\}$ **do**
11. Fit K-Means model with k clusters
12. Compute Within-Cluster Sum of Squares W_k
13. Store W_k
14. **end for**
15. Plot WCSS curve
16. Identify optimal k using elbow point
17. Apply K-Means on dataset with optimal k
18. Assign cluster labels C to dataset D

Supplementary Table 4:
Notations used in Algorithm 1

Notation	Description
D	Dataset with n rows and m columns
K	Range of clusters
k	Number of clusters
S_k	Silhouette score for k clusters
W_k	Within-cluster sum of squares
C_j	Cluster j in dataset
μ_j	Centroid of cluster C_j
x_i	Datapoint i
C	Final cluster assignments
r	Random seed

Macro-Averaged Precision:

$$\text{Precision}_{macro} = \frac{1}{C}\sum_{i=1}^{C}\text{Precision}_i$$

Weighted Precision:

$$\text{Precision}_{weighted} = \sum_{i=1}^{C} w_i \cdot \text{Precision}_i \quad \text{where } w_i = \frac{\text{Support}_i}{\text{Total Samples}}$$

3. Recall

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i}$$

Macro Recall:

$$\text{Recall}_{macro} = \frac{1}{C}\sum_{i=1}^{C}\text{Recall}_i$$

Weighted Recall:

$$\text{Recall}_{weighted} = \sum_{i=1}^{C} w_i \cdot \text{Recall}_i$$

4. F1-Score

$$F1_i = 2 \cdot \frac{\text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}$$

Macro F1:

$$F1_{macro} = \frac{1}{C}\sum_{i=1}^{C} F1_i$$

Weighted F1:

$$F1_{weighted} = \sum_{i=1}^{C} w_i \cdot F1_i$$

5. Validation Loss

The categorical cross-entropy loss is used for multi-class classification:

$$\mathscr{L} = -\frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{C} y_{n,i} \log(\hat{y}_{n,i})$$

where:

- N = number of samples
- C = number of classes
- $y_{n,i}$ = 1 if sample n belongs to class i, else 0
- $\hat{y}_{n,i}$ = predicted probability for class i

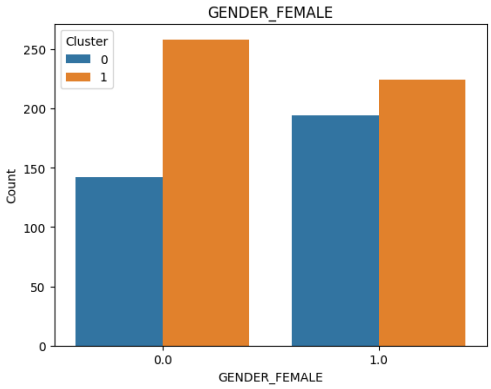
Averaging Strategy:

- *Macro Average*: Treats all classes equally.
- *Weighted Average*: Weights each class by its number of samples.

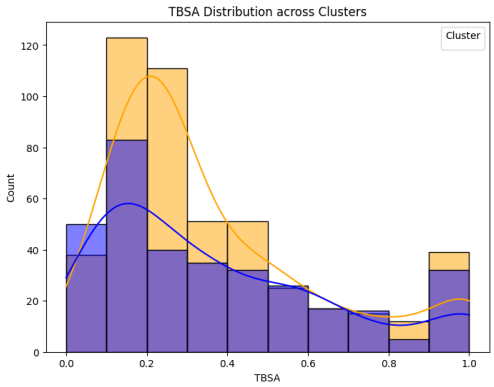
3 Clustering Patient's Data

Cluster Characteristics and Distribution

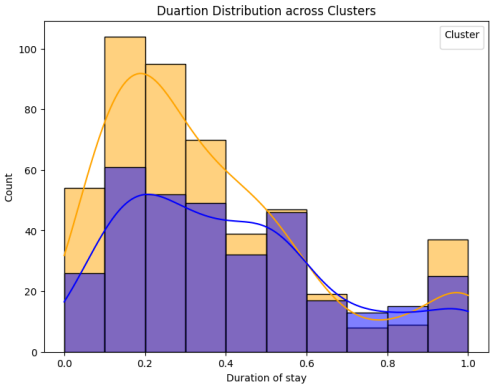
After implementing K-Means clustering, cluster labels are assigned to dataset so we can analyze patients' characteristics and distribution across the clusters.



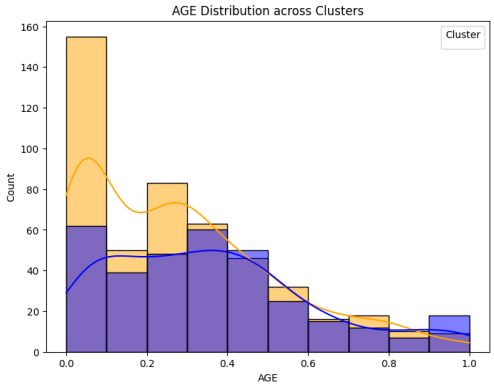
(a) Gender distribution across each cluster



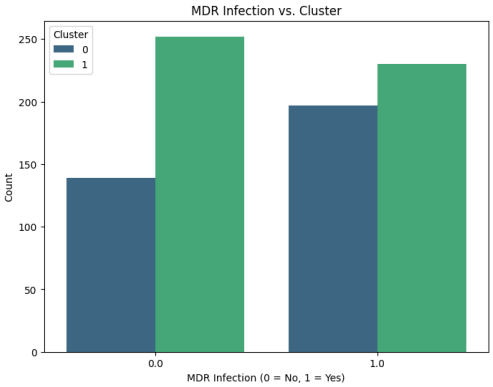
(b) TBSA across each cluster



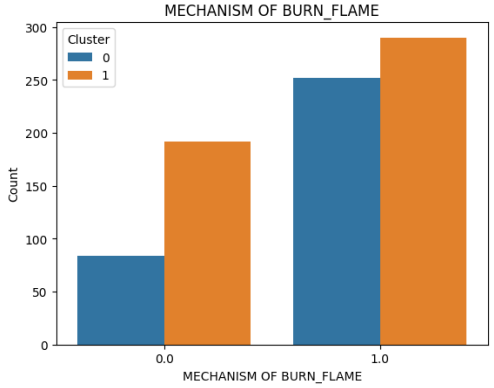
(c) Duration of stay distribution across each cluster



(d) Age distribution across each cluster



(e) MDR infection across each cluster

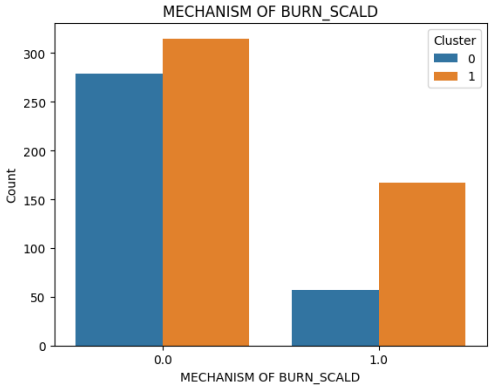


(f) Mechanism of burn flame across each cluster

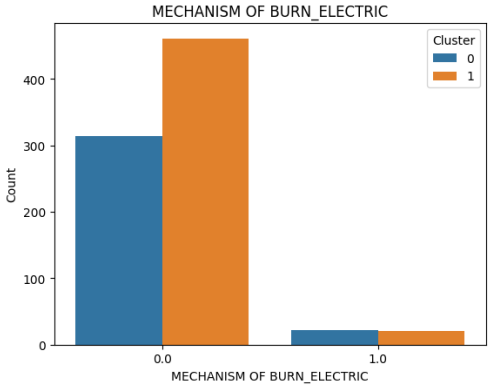
Supplementary Figure 6: Patient characteristics across clusters. The figure shows distributions of gender, TBSA, duration of stay, age, MDR infection, and burn mechanism across the identified patient clusters.

Figure 6a shows cluster 1 has greater number of both males and females than Cluster 0. Both genders are present in both clusters, but females dominate slightly in Cluster 1. TBSA values are right-skewed (most patients have lower TBSA values). Figure 6b shows Cluster 1 has more patients with slightly higher TBSA values compared to Cluster 0. Figure 6c shows that cluster 1 has more patients with longer durations of stay, suggesting that patients in Cluster 1 may have more severe burns or complications. Figure 6d shows cluster 0 has more patients at a younger age range, while Cluster 1 is slightly more spread across the age range, possibly indicating older or more diverse patient ages. Figure 6e shows MDR infections are more prevalent in Cluster 1 than in Cluster 0, suggesting Cluster 1 may include more critical cases or longer stays (increasing infection risk). Figure 6f shows that flame burns are more common

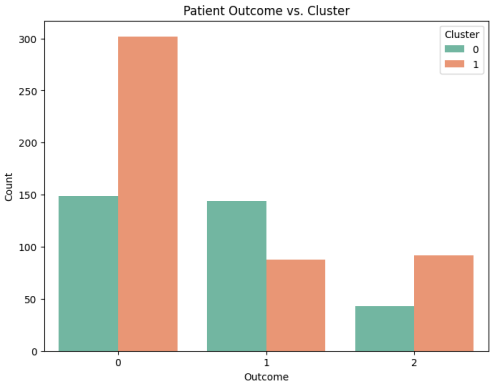
in Cluster 1, while Cluster 0 has a mix of flame and other mechanisms. This implies Cluster 1 might represent more severe or specific types of burn cases (like flame burns).



(a) Mechanism of burn scald distribution across each cluster



(b) Mechanism of burn electric across each cluster



(c) Outcome distribution across each cluster

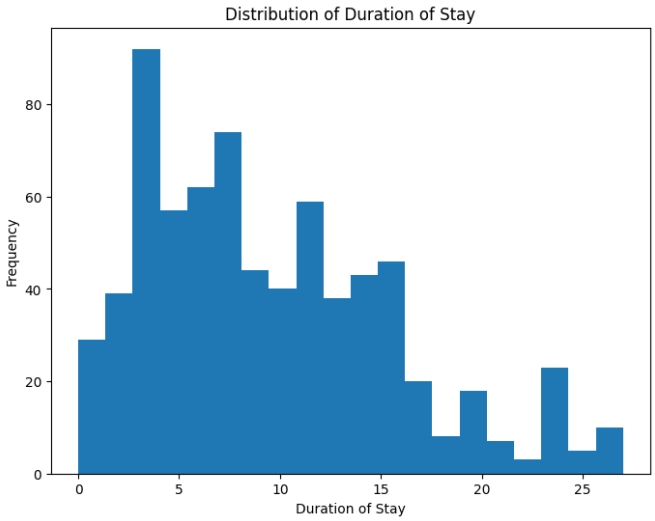
(d) Depth of burn across each cluster

Supplementary Figure 7: Burn mechanisms and outcomes across clusters. The figure shows distributions of scald and electric burn mechanisms, patient outcomes, and burn depth across the identified patient clusters.

Figure 7a illustrates scald burns are more common in cluster 1. Cluster 0 has more non-scald burns. Suggests Cluster 1 may represent patients with less severe burn types (scalds often affect younger individuals or are domestic). Figure 7b illustrates electric burns are rare but slightly more in cluster 0. Cluster 1 has very few electric burn cases. May suggest cluster 0 captures more high-risk or occupational burns (electric burns often being more severe and work-related). Figure 7c shows cluster 1 has more discharged patients. cluster 0 has a higher count for LAMA and Death, especially death. This suggests that Cluster 0 represents more severe or complicated cases, possibly with worse prognosis. Figure 7d illustrates cluster 0 has higher counts for deeper burns. Cluster 1 dominates at lower depth levels.

Forecasting Duration of Stay

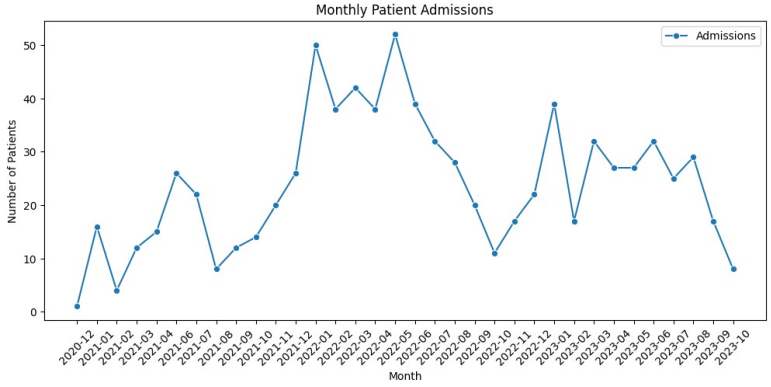
Prior to using feed forward Neural Network, trends in data has been discovered through data visualization. As distribution is right skewed, some patients needed longer hospital stays and majority need short stays.



Supplementary Figure 8: Distribution of Duration of Stay

Figure 8 is a histogram plot showing the distribution of the duration of stay for patients with burns. The length of stay (days) plotted on the x-axis represents the length of stay in hospital, and the frequency of patients staying for that length on the y-axis. Histogram is right-skewed, meaning that most patients have a shorter period of hospitalization and few patients are continuing to stay longer. Peaks in duration of stays of 4-6 days and 10-12 days coincide with the most common periods of hospitalization for burns.

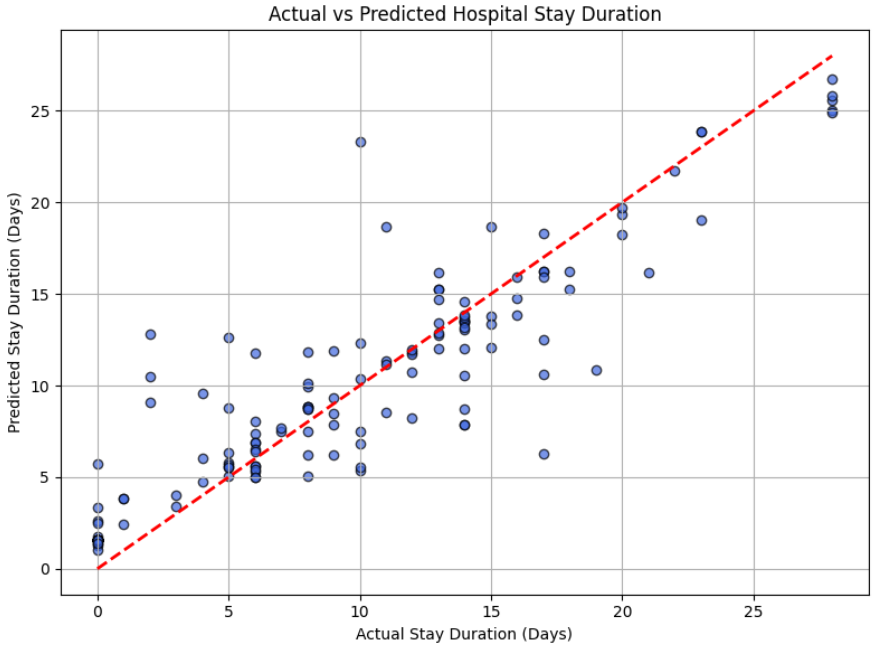
Longer hospital stays are typically seen in patients with higher TBSA. Longer stays and need for more intense treatment are associated with deeper burns. Winter burn admissions are higher, as a result of incidents involving flame and scald. There is no distinction in the duration of stay for male and female patients.



Supplementary Figure 9: Seasonal trends

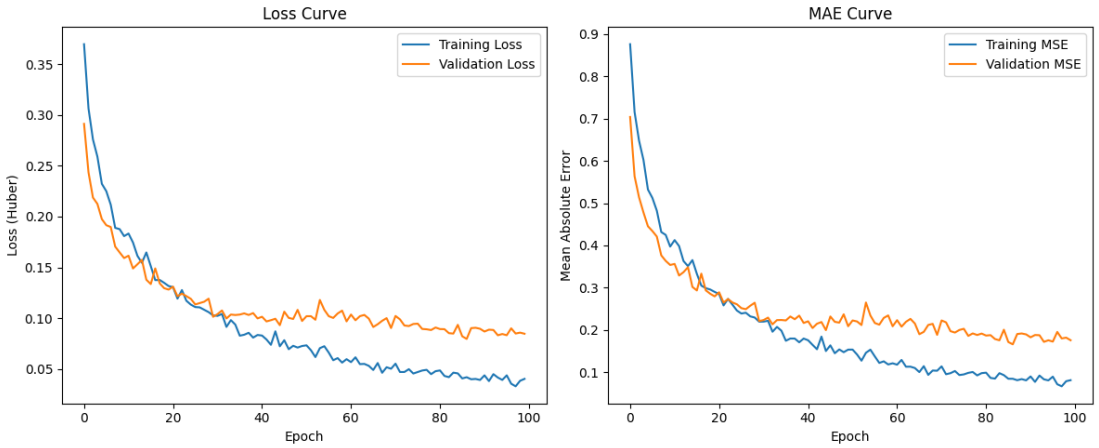
The line graph in Figure 9 is representing monthly patient admissions over a period of time. Months are shown in calendar month (from December 2020 to October 2023). Percentage of the admitted patients is shown on the y-axis. The trends indicate seasonal variations as peak month occurrences occur in winter months (December 2021, May 2022, December 2022) and decline month occurrences occur in summer months (April–May 2022), followed by a steady decline in that season.

4 Predicted Model Performance Analysis



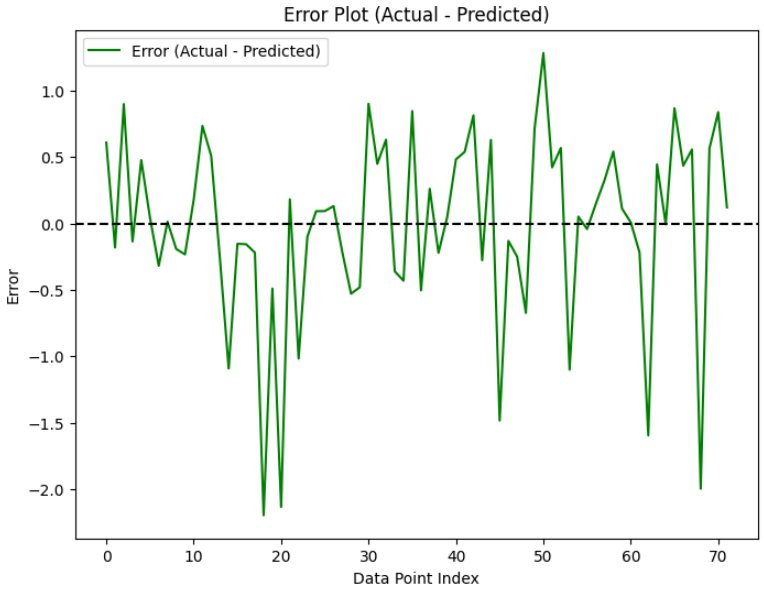
Supplementary Figure 10: Actual vs Predicted scatter plot

In Figure 10 each blue dot represents a data point, with the x-axis indicating the actual number of days a patient stayed and the y-axis showing the corresponding model prediction. The red dashed diagonal line represents the ideal scenario where the predicted stay exactly matches the actual stay (i.e., perfect predictions). The proximity of most points to this line suggests that the model performs reasonably well.



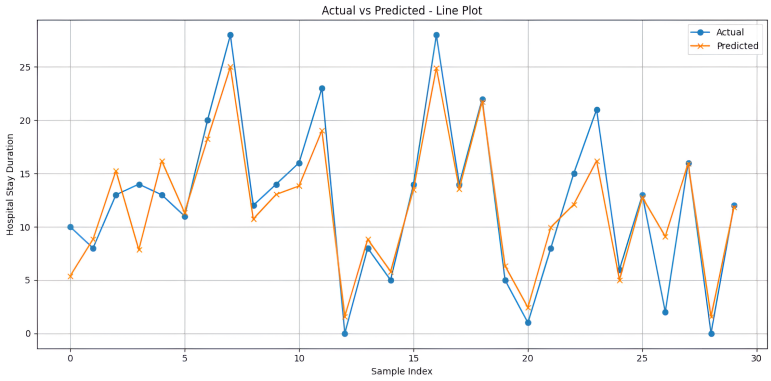
Supplementary Figure 11: ANN Training Loss Plot

Figure 11 shows training vs. validation loss over an epoch interval during model training. The x-axis is the number of epochs where the y-axis is the loss. At initial training loss (blue) and validation loss (orange) are high, but quickly decrease as a sign that the model is learning effectively. After a few epochs the loss stabilizes at near zero, which is a sign of good convergence, without overfitting. An equilateral distribution of the training and validation loss curves also implies good generalization to unknown data, which are desirable for predictions.



Supplementary Figure 12: Error distribution plot

In Figure 12 the plot describes the difference between actual and fore-casted values. Errors are normally distributed, illustrating no systematic biases in predictions.



Supplementary Figure 13: Actual vs Predicted Line plot

As shown in Figure 13 the model provides reasonable predictions, as demonstrated by the close correspondence between expected values and actual values.

References

[1] DP Yadav, Turki Aljrees, Deepak Kumar, Ankit Kumar, Kamred Udham Singh, and Teekam Singh. Spatial attention-based residual network for human burn identification and classification. *Scientific Reports*, 13(1):12516, 2023.

[2] Yong-Zhen Huang, Yan-Ming Chen, Chih-Cheng Lin, Hsiao-Yean Chiu, and Yung-Chun Chang. A nursing note-aware deep neural network for predicting mortality risk after hospital discharge. *International Journal of Nursing Studies*, 156:104797, 2024.

[3] Samantha Huang, Justin Dang, Clifford C Sheckter, Haig A Yenikomshian, and Justin Gillenwater. A systematic review of machine learning and automation in burn wound evaluation: A promising but developing frontier. *Burns*, 47(8):1691–1704, 2021.

[4] Justin J Lee, Mahla Abdolahnejad, Alexander Morzycki, Tara Freeman, Hannah Chan, Collin Hong, Rakesh Joshi, and Joshua N Wong. Comparing artificial intelligence guided image assessment to current methods of burn assessment. *Journal of Burn Care & Research*, 46(1):6–13, 2025.

[5] Mahmoud E Khani, Zachery B Harris, Omar B Osman, Juin W Zhou, Andrew Chen, Adam J Singer, and M Hassan Arbab. Supervised machine learning for automatic classification of in vivo scald and contact burn injuries using the terahertz portable handheld spectral reflection (phasr) scanner. *Scientific Reports*, 12(1):5096, 2022.

[6] D. P. Yadav, A. S. Jalal, and Ved Prakash. Human burn depth and grafting prognosis using resnext topology based deep learning network. *Multimedia Tools and Applications*, 81:18897 – 18914, 2022.

[7] Joohi Chauhan and Puneet Goyal. Convolution neural network for effective burn region segmentation of color images. *Burns*, 47(4):854–862, 2021.

[8] DM Anisuzzaman, Chuanbo Wang, Behrouz Rostami, Sandeep Gopalakrishnan, Jeffrey Niezgoda, and Zeyun Yu. Image-based artificial intelligence in wound assessment: a systematic review. *Advances in wound care*, 11(12):687–709, 2022.

[9] Metin Yıldız, Yakup Sarpdağı, Mehmet Okuyar, Mehmet Yildiz, Necmettin Çiftci, Ayşe Elkoca, Mehmet Salih Yildirim, Muhammet Ali Aydin, Mehmet Parlak, and Bünyamin Bingöl. Segmentation and classification of skin burn images with artificial intelligence: Development of a mobile application. *burns*, 50(4):966–979, 2024.

[10] Sirisha Rambhatla, Samantha Huang, Loc Trinh, Mengfei Zhang, Boyuan Long, Mingtao Dong, Vyom Unadkat, Haig A Yenikomshian, Justin Gillenwater, and Yan Liu. Dl4burn: Burn surgical candidacy prediction using multimodal deep learning. In *AMIA Annual Symposium Proceedings*, volume 2021, page 1039, 2022.

[11] Che Wei Chang, Chun Yee Ho, Feipei Lai, Mesakh Christian, Shih Chen Huang, Dun Hao Chang, and Yo Shen Chen. Application of multiple deep learning models for automatic burn wound assessment. *Burns*, 49(5):1039–1051, 2023.

[12] Muhammad Ashraf, Hafiz MuhammadAfzaal, Muhammad Azam, Tahir Akram, et al. Aa study on the effectiveness of fine-tuning vgg16 versus classical machine learning and deep learning models for skin burn detection. *Journal of Innovative Computing and Emerging Technologies*, 4(2), 2024.

[13] C Pabitha and B Vanathi. Densemask rcnn: A hybrid model for skin burn image classification and severity grading. *Neural Processing Letters*, 53(1):319–337, 2021.

[14] Ji Hyun Park, Yongwon Cho, Donghyeok Shin, and Seong-Soo Choi. Prediction of mortality after burn surgery in critically ill burn patients using machine learning models. *Journal of Personalized Medicine*, 12(8):1293, 2022.

[15] Saeka Rahman, Miad Faezipour, Guilherme Aramizo Ribeiro, Elika Ridelman, Justin D Klein, Beth A Angst, Christina M Shanti, and Mo Rastgaar. Inflammation assessment of burn wound with deep learning. In *2022 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 1792–1795. IEEE, 2022.