Generative AI for Recognizing Nurse Training Activities in Skeleton-Based Video Data

Md Ibrahim Mamun* [1], Shahera Hossain [2], Md Baharul Islam [3], Md Atiqur Rahman Ahad [4]
[124]University of East London, [3]Florida Gulf Coast University

Abstract

Endotracheal suctioning (ES) is a complex procedure associated with a series of actions and inherent risks, particularly in the intensive care unit (ICU). Given the importance of precise execution, it is preferable to have skilled nurses perform ES tasks. To facilitate nurse training and ensure proficiency in ES procedures, automated nursing activity recognition presents a promising solution, offering benefits in terms of cost, time, and effort. In this paper, we propose a novel approach to nurse training activity recognition for ES tasks, leveraging the capabilities of Generative Artificial Intelligence (GenAI). Specifically, we demonstrate how Large Language Models (LLMs), a subset of GenAI, can enhance the efficiency of nursing activity recognition. By employing LLMs such as OpenAI's Generative Pre-trained Transformer (ChatGPT), Google's Gemini, and Microsoft's Copilot, we aim to improve the accuracy and efficiency of our methodology. Our study identifies a clear gap in the utilization of LLMs for more accurate determination of nursing activities related to ES, with reduced human interaction. Through the integration of approaches and data features suggested by LLMs, we achieve a notable increase in accuracy from baseline 0.51 to 0.58, along with an elevated F1 score from 0.31 to 0.46. These results underscore the potential of LLMs, as a subset of GenAI, to enhance traditional problem-solving efficiency by offering robust solutions and procedures.

1 Introduction

The field of computer vision (CV) is giving a significant focus on Human Activity Recognition (HAR) which refers to the prediction of hu-

[1]mamun.cs104@gmail.com
[2]shaherahossain@gmail.com
[3]mislam@fgcu.edu
[4]mahad@uel.ac.uk

man body motion and movement using data generated by various sensors and devices[19]. There are plenty of applications and effects of analyzing HAR. Recognizing daily life activities is becoming increasingly valuable for enhancing efficiency, reducing costs, and streamlining efforts which is particularly important for applications in healthcare, education, and manufacturing industries, which involve abnormal behavior detection and human-computer interaction[18]. In this paper, we analyze and predict nursing activity recognition regarding identifying and predicting tasks associated with ES. ES is one of the critical activities that requires experienced nurses. To train a nurse and examine their performance, HAR can be a beneficial tool in different ways by minimizing time and cost.

GenAI and LLMs are technological wonders known for their power in natural language processing and content generation which creates numerous opportunities and ways to solve real-world problems[6]. GenAI and LLMs are very useful in revolutionizing data and information management in various sectors including healthcare, medicine, production, and more with their unique ability to generate new data or contents, providing promising guidance and solutions[22]. In this paper, we show how LLMs influenced the performance of our deployed methods in nursing activity recognition regarding ES.

Here we aim to show how a subset of GenAI that is LLMs helps to enhance the performance of our selected models by demonstrating for a particular dataset, how recommendations and solutions generated by ChatGPT 4.0, ChatGPT 3.5, Gemini, and Copilot which are the products of modern LLMs increases both accuracy and F1 score in our methodology. In a broader sense, GenAI holds the capability to transform healthcare applications, predictions, and data management through automated procedures, increasing clinical decision-making and offering expert diagnostic supports and tools[17]. Table 1 shows some of the fundamental differences between the traditional approach and the GenAI approach.

The structure of our methodology is divided into two sections where the first part consists of a traditional approach with a minor influence of LLMs. In the second part of the methodology all features, algorithms, and processes are followed and used according to LLMs. After the successful completion of the LLMs part, we achieved better accuracy than the traditional approach for the training dataset. The demonstration not only showcases promising outcomes but also unveils opportunities and insights that surpass human capabilities, all within a specific timeframe. GenAI holds unique capabilities to modify existing content and processes, create new data and processes, and represent a revolutionary transformation compared to adding technological addition[10].

This paper is divided into the following sections. Section 2 discusses a literature review focusing on both recent and past papers related to nursing activity recognition through skeleton-based video data focusing on LLMs. Section 3, shows data description, data preparation, and data pre-processing. Section 4 shows our proposed Methodology, Section 5 shows the result analysis. In section 6 we discuss the advantages and disadvantages of our proposed methodology of using LLMs, a subset of GenAI. Finally, Section 7 concludes the paper.

Table 1: Overview of Traditional and GenAI Interactive Approaches

Approach	**Traditional**	**GenAI Interactive**
Manual	✓	✓
Rule-based	✓	✓
Linear	✓	✓
Structured	✓	✓
Deterministic	✓	✓
Repetitive	✓	✓
Human Intervention	✓	✓
Historical-data	✓	✓
Semantic Analysis	-	✓
Automated	-	✓
Data-driven	-	✓
Dynamic Adaption	-	✓
Innovative	-	✓
Scalable	-	✓
AI-driven Insight	-	✓
Self-learning	-	✓
Customizable Algorithms	-	✓
Probabilistic	-	✓
Real-Time Analysis	-	✓

2 Literature Review

This section discusses both present and past works related to nursing activity recognition mainly from skeleton-based video data focusing on LLMs, a subset of GenAI.

In [20], authors proposed a feature selection strategy to select only important features from the handcrafted features. They also claimed that their simple classifier can generate satisfactory performance on the nurse care activity data. However, we did not find any activity recognition related to ES or any feature selection to recognize ES activities. Also, the authors mentioned only the use of the Spatio-temporal graph convolution method in their methodology. We did not find any use of LLMs that is GenAI, especially during the feature selection process. On the other hand, we have used LLMs-generated processes and features to recognize nursing activities related to ES which is a critical task nurses need to perform in ICU when necessary.

In [4], authors mentioned ES as a life-threatening technique and discussed the importance of correct and immediate actions to complete ES tasks. Authors mainly focused on the prevention of secondary infection by suctioning which is one of the activities in ES. However, we did not find any actions toward recognizing whether the activities during the ES procedure were correctly followed or not. In this paper, we aim to use skeleton-based data to recognize whether nurses are correctly performing tasks in the ES process. Therefore, to recognize ES activities more precisely and aid to nurse training in ES our methodology aims to enhance accuracy and effectiveness through the use of LLMs.

In [2] [15] [9], authors showed a different approach to recognizing human activity mainly using skeleton data. Here author used a probabilistic graphical model, skeleton-based spatiotemporal informational and conventional neural networks (CNN), and a method composing 3D skeleton data to recognize human everyday activities and classification of human activities. However, the authors did not mention any activity recognition methods that can aid nurse training focusing on ES activity recognition. Also, the authors mainly used traditional AI approaches in their proposed solutions. The authors did not provide any specific opportunities and challenges regarding the use of LLMs.

From [14] [13], we got our motivation to recognize nurse training activity in ES using video-based pose estimation data i.e. skeleton data. We have used their dataset to build and use our methodology. Here authors mentioned a structured way to recognize nursing activities in ES. The authors used You Only Look Once version 7 (YOLOv7) to extract pose estimation from video data. The authors represented the pose estimation output from each frame using YOLOv7 through a vector comprising X, and Y coordinates and the confidence (c) of the keypoints as illustrated by equation 1.

$$F = [x_0, y_0, c_0, x_1, y_1, c_1, \dots, x_{16}, y_{16}, c_{16}] \tag{1}$$

The authors adeptly managed scenarios involving the presence of two or more individuals in a frame. They approached this by calculating the trunk length according to equation 2, considering that the nurse standing near the camera is the one who is carrying out ES tasks, and the subject's trunk length is likely to be greater than that of any other individual in the background.

$$\begin{multline} \text{trunk length} = 1.2 \times \text{left forearm length} \\ + \text{left upper arm length} + 0.1 \times \text{hip length} \\ + 0.1 \times \text{shoulder length} \tag{2} \end{multline}$$

Here authors used two approaches which are the Baseline approach and our proposed approach. However, we did not find any clear discussion about using LLMs such as ChatGPT, Gemini or Copilot to enhance the proposed solution. Also, the authors did not show any comparison between two or more LLMs for improvement of any specific part of the problem. In this paper, we aim to use popular LLMs through two-way communication to enhance the solutions for nursing activity recognition in the realm of ES. Moreover, we aim to show a comparison between different LLMs so that valuable insight can be extracted for specific parts of the problem.

In [11] [7], the authors extracted features from sensor data, then used K-Nearest Neighbors (KNN) and achieved highest accuracy on 10-fold cross-validation. Their dataset contains nurse activity data and was collected using three sensors named Motion Capture, Meditag, and Accelerometer sensors. However, we did not find the use of any other algorithms. In our methodology, we have used various types of algorithms to analyze which performs better in our dataset.

In [1] [5], authors used 3D pose estimation and multilabel classification to improve human activity recognition focusing on nursing activity as well. In both of the cases, authors achieved 55.7% and 63% F1 Score. However, we did not find any scope for using LLMs to enhance feature selection and recommendations for selecting algorithms through two-way communication.

In [3], the authors proposed a spatial-temporal graph convolutional network (ST-GCN) to recognize nurse care activity. Their model was evaluated by using leave-one-subject-out cross-validation methods and achieved 57% accuracy. However, we did not find any recommendations regarding the use of other close algorithms such as Temporal Convolutional Networks or Light Gradient Boosting Machines. In our research work, we used both mentioned algorithms to analyse and get valuable insight regarding the performance of our model.

In [8], the authors proposed a novel method named LLMIE-UHAR which leverages LLMs and Iterative Evolution to realize Unsupervised HAR. They used LLMs to fuse both contextual and semantic information and annotate key samples through a clustering algorithm. The authors also used multiple approaches to analyze outcomes using different models. However, the authors used data which is based home home activities such as reading, and other household activities. However, the author did not specify whether their proposed model, LLMIE-UHAR, is suitable for recognizing nurse training activities that focus on ES. In environments like nurse training activities focusing on ES, there are typically fixed activities that occur in a specific sequence. Additionally, there may be two or more nurses in the background. Conversely, in a home environment, activities can vary widely and do not require sequencing.

3 Data Preparation

3.1 Data Description

Pose Skeleton (Keypoints) data are extracted from video data recorded from ten nurses and twelve nursing students performing ES on the ESTE-SIM simulation. Here YOLOv7 was used to extract pose skeleton data from the videos. There are a total of 9 activities in the ES procedure. Figure 1 shows the joint angles which are used in feature extraction.

Table 2: ES Activities and their ID's

Activity Class ID	Activity Name
0	Catheter preparation
1	Temporal removal of an artificial airway
2	Suctioning phlegm
3	Refitting the artificial airway
4	Catheter disinfection
5	Discarding gloves
6	Positioning
7	Auscultation
8	Others

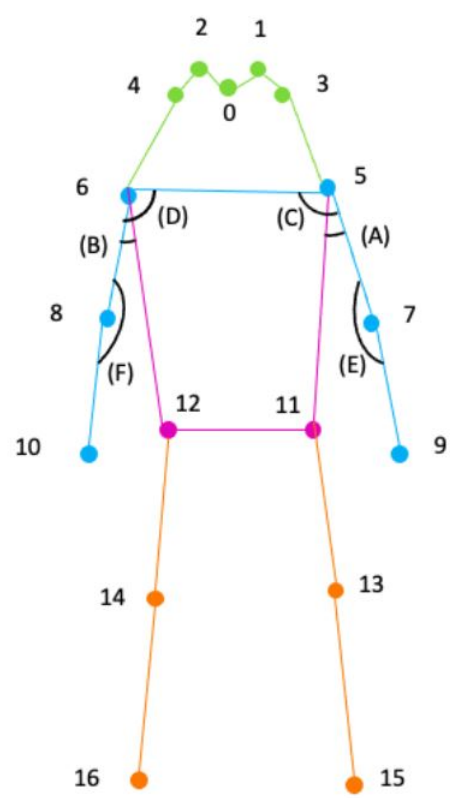

Figure 1: Joint angles A,B,C,D,E are used in feature extraction[14].

3.2 Data Structure

There are also 32 CSV files containing annotations of each video in the training set. Annotation data contain the start time, stop time, activity

Table 3: Keypoint Data Description

Keypoint	**Coordinates (X, Y)**	**Confidence**
Nose	nose_x, nose_y	nose_conf
Eye (L/R)	left_eye_x, left_eye_y	left_eye_conf
	right_eye_x, right_eye_y	right_eye_conf
Ear (L/R)	left_ear_x, left_ear_y	left_ear_conf
	right_ear_x, right_ear_y	right_ear_conf
Shoulder (L/R)	left_shoulder_x, left_shoulder_y	left_shoulder_conf
	right_shoulder_x, right_shoulder_y	right_shoulder_conf
Elbow (L/R)	left_elbow_x, left_elbow_y	left_elbow_conf
	right_elbow_x, right_elbow_y	right_elbow_conf
Wrist (L/R)	left_wrist_x, left_wrist_y	left_wrist_conf
	right_wrist_x, right_wrist_y	right_wrist_conf
Hip (L/R)	left_hip_x, left_hip_y	left_hip_conf
	right_hip_x, right_hip_y	right_hip_conf
Knee (L/R)	left_knee_x, left_knee_y	left_knee_conf
	right_knee_x, right_knee_y	right_knee_conf
Ankle (L/R)	left_ankle_x, left_ankle_y	left_ankle_conf
	right_ankle_x, right_ankle_y	right_ankle_conf

name, activity ID, and more.

Table 4: Annotation Data Description

Column Name	**Description**
start_time	Start time of activity (in second)
stop_time	Stop time of activity (in second)
annotation_str	Activity name
annotation	Activity ID

Here Table 2 shows the activities in ES and their ID. In the activity name, we have Catheter preparation, Temporal removal of an artificial airway, Suctioning phlegm, Refitting the artificial airway, Catheter disinfection, Discarding gloves, Positioning, Auscultation, and Others. Table 3 shows the key point data description. keypoint data include the nose, eyes, ears, shoulders, elbows, wrists, hips, knees, and ankles are identified with specific coordinates (X, Y) and associated confidence levels for each. Table 4 shows the annotation data description as mentioned earlier.

4 Methodology

In this paper, we propose to use LLMs, subset of GenAI to build a model that comprises ML feature engineering and ML algorithms to recognize nursing activity regarding ES. We have divided our methodology into two parts. The first part contains basic data processing including removing redundant data and removing features such as knee and ankle coordinates and their confidences which may not be important to recognize activity. Here some of the parts are added according to the generated suggestions of LLMs such as ChatGPT, Gemini and Copilot. The second part contains more robust suggestions regarding new features that can be calculated from Skeleton-based data, and different algorithms that may increase the efficiency of recognizing nursing activity.

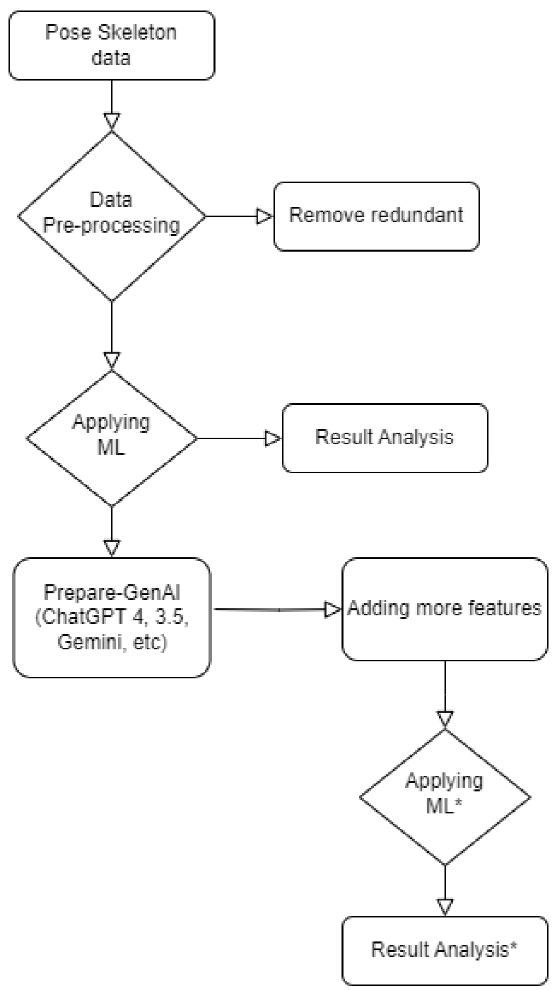

Figure 2: A concise flow chart of the overall work process.

According to the flow-chart of Figure 2, pose skeleton data are going through initial data reprocessing. Here some of the basic feature engineering is done regarding removing redundant data and more. The ML algorithm is used and results are obtained.

$$\mathrm{PCA(X)} = \mathrm{XW} \tag{3}$$

Here X represents the original data matrix, and W represents the matrix of principal components.

Then with the help of LLMs, new features are added and various algorithms are used to identify the best-performing algorithm for this specific problem, all are done mostly through two-way communication with LLMs tools such as ChatGPT 4, Gemini, and Copilot.

In our methodology we used confusion matrix which is a popular tool used in classification tasks to evaluate the performance of a classification model.

$$\text{Objective function} = \sum_{i=1}^{n} \mathrm{loss}(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{4}$$

Here l is the loss function, $\hat{y}_i(t)$ is the predicted value at iteration t, and $\Omega(f_k)$ is the regularization term for the k-th tree.

$$\text{Objective Function} = \sum_{i=1}^{n} l(y_i, \hat{y}_i(t)) + \sum_{k=1}^{K} \Omega(f_k) \tag{5}$$

Here loss is the loss function, $\hat{y}_i$ is the predicted value, and $\Omega(f_k)$ is the regularization term for the k-th tree.

Here Table 5 shows the algorithm overview of the methodology. On the other hand, Table 6 shows one of the refined versions of communication. Here, we have demonstrated one of the communications which are done in terms of asking which features can be calculated given the problem statement and data context, and which ML algorithms we should use after calculating some of the suggested features.

It is visible that LLMs can suggest improvement and experiment analysis according to the given context such as an expert in this domain. Here, ML is given priority as it is more popular in the domain of healthcare and activity recognition [12] [21]. Equation (3), (4), and (5) represents Principal Component Analysis, XGBoost algorithm, and LightGBM algorithm respectively.

Table 5: Nursing Activity Recognition Algorithm Overview using LLMs

Stage	Description
Input	Pose skeleton data from video sequences, annotation data for activities.
Output	Activity type predictions for each video segment with improved accuracy.
1. Initial Method	Utilized basic pose skeleton data extraction and statistical feature extraction leading to lower accuracy in activity predictions.
2. Data Preprocessing	Applied segmentation and smoothing on pose skeleton data to prepare for enhanced feature extraction.
3. Initial Feature Extraction	Extracted basic features such as mean, standard deviation, and velocities which resulted in lower model performance.
4. Initial Model Evaluation	Initial model training with traditional machine learning techniques showed sub-optimal results and accuracy.
Introduction of Generative AI	**Incorporated Generative AI techniques to significantly enhance feature extraction and model performance.**
5. Enhanced Feature Extraction using LLMs (GenAI)	- Introduced angular features between joints and trajectory analysis for dynamic movement capture. - Calculated advanced features like joint angles, velocity changes, and trajectory curvature using Generative AI.
6. Model Training with LLMs	Re-trained the model using LightGBM with enhanced features, leading to significantly improved performance.
7. Enhanced Model Evaluation	The application of Generative AI techniques resulted in a marked increase in F1 score, accuracy, precision, and recall.
8. Activity Prediction with LLMs	Achieved higher prediction accuracy for unseen video segments, demonstrating the effectiveness of Generative AI enhancements.
9. Results Visualization	Visualized the improved keypoint trajectories and model predictions, showcasing the superior performance of the enhanced model.

Table 6: Two-way Communication with LLMs and Extended Responses to Queries

GenAI Models	**Refined Q1: What key features can be identified from skeleton data for endotracheal suctioning activity recognition?**	**Refined Q2: Post-PCA, which algorithms enhance accuracy and F1 score?**
ChatGPT (Analytics Features Precision Recognition Recommendations)	ChatGPT leverages advanced analytics on angular measurements, speed, and path of limb movements, aiming to precisely identify the execution of medical tasks. This approach is key in separating similar tasks through nuanced differences in motion, providing an accurate basis for recognizing specific nursing actions such as endotracheal suctioning.	Recommends LightGBM for its efficiency with large datasets and complex features, alongside XGBoost, Random Forest, SVM, and CNNs for their robust performance across various data types.
Gemini (Posture Moments Phase Recommendations)	Gemini's methodology involves a deep dive into the posture analysis, detailed arm movement tracking, and the interactions between the nurse and medical equipment. By focusing on how these elements change throughout the suctioning process, Gemini offers insights into identifying critical moments and ensuring accurate phase recognition, crucial for training and performance assessment.	Suggests SVM for high-dimensional spaces, Random Forest for its robustness to outliers, and GBC for handling complex classification problems. Emphasizes the importance of CNNs for large data sets.
Copilot (Integration Nuanced Recognition Highlights)	Copilot employs a comprehensive framework that integrates pose estimation with analysis of joint angles and sequential movement patterns. This allows for an in-depth understanding of the nurse's actions, from preparation to the actual suctioning and completion, offering a nuanced view that supports both activity recognition and ergonomic assessments. Such detailed analysis is essential for developing targeted training programs and improving nursing care techniques.	Highlights Logistic Regression for binary classification, Random Forest for non-linear relationships, SVM for optimal hyperplane separation, Neural Networks for complex relationships, and Ensemble Methods for improved performance.

We have used multiple algorithms suggested by GenAI, to check which algorithms perform best in the training data. In the first part before calculating features suggested by GenAI we used eXtreme Gradient Boosting (XGBoost), Convolutional neural network (CNN), and Long short-term memory (LSTM) which are also outcomes of two-way communication with LLMs. On the other hand, in the second part, we calculated the Angular Feature, Velocity Acceleration, Trajectory Features, Temporal Features, Energy Features, and Statistical Moments. However some of the features were causing poor accuracy and F1 score, therefore we finalize Angular Feature, Velocity Acceleration, and Trajectory Features. Also, we have tried different algorithms such as light gradient-boosting machine (LightGBM), TAB Transformer, Feedforward Neural Network (FNN), CATBoost, Entity Embedding, Temporal Convolutional Networks (TCN), Decision Tree (DT), Support Vector Machine (SVM), Random Forest Classification. We also used a hybrid model consisting of CNN or GCN, LSTM/GRU, and LightGBM. Moreover, we tried to use Vision Transformer (ViT) using GenAI, but due to a lack of Random Access Memory (RAM), the program crashed in Google Colab.

5 Result Analysis

In this section, we aim to analyze the results before and after using LLMs' suggested development and solutions in our methodology. As mentioned in the methodology section, we calculated various features and only finalized those features which contributed to better accuracy and F1 Score. Similarly, in the second part, we implemented and used different algorithms and approaches and only selected the best-performing algorithm in terms of the highest accuracy and F1 Score.

It is worth mentioning that we also checked feature importance in our training dataset and tried to make our model better focusing on the highest feature value. But in every case, F1 and accuracy were slightly decreasing.

In the first section of the program, we finalized using XGBoost algorithms and got an F1 Score of 0.31 and an Accuracy of 0.51 with a Precision of 0.38 and a Recall of 0.30. Figure 4 shows the confusion matrix (CM) generated after using XGBoost.

In the second section after calculating and selecting new features suggested by different LLMs, we finalized using the LightGBM algorithm and achieved a higher Accuracy of 0.58 and an F1 score of 0.46 with a Precision of 0.61 and a Recall of 0.41. Figure 5 shows CM for the second part of the program.

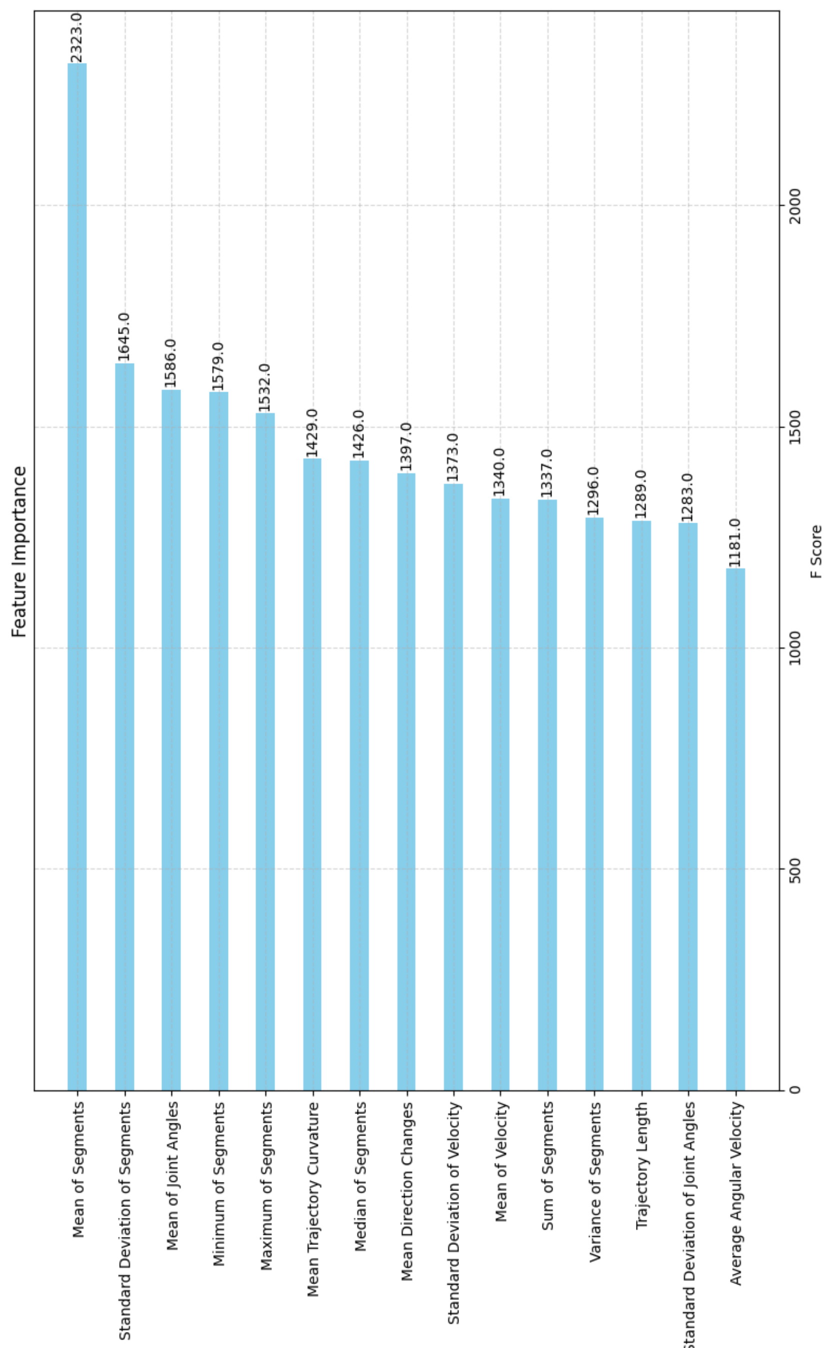

Figure 3: Feature Importance Analysis.

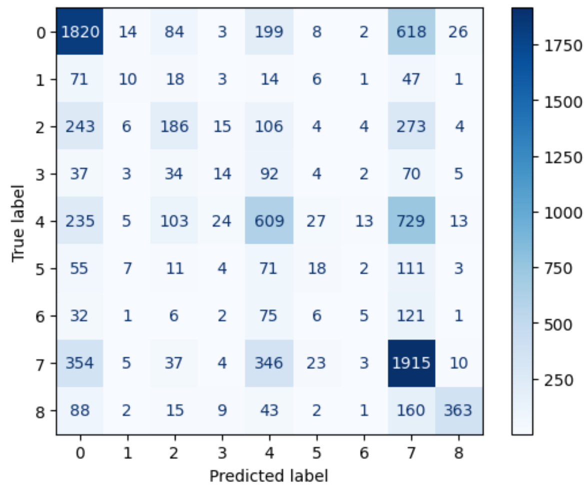

Figure 4: Confusion Matrix for first section.

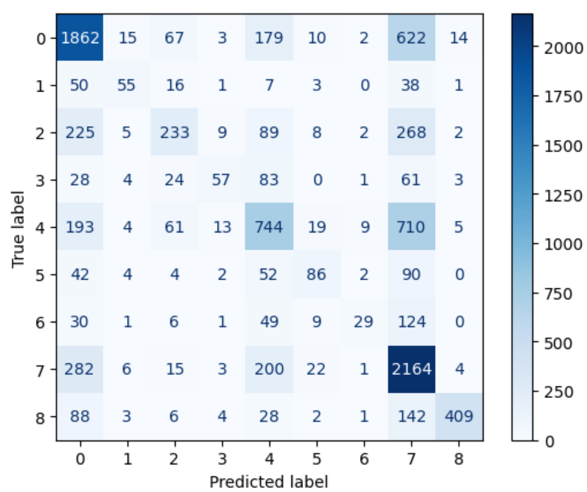

Figure 5: Confusion Matrix for second section.

Table 7 shows the feature calculation suggested by LLMs, and after using specific ML algorithms on those features, calculates the F1 Score and Accuracy of the model. To make the program more efficient, we finally selected three features - Angular Feature, Velocity Acceleration, and Trajectory Features. We also identified the most contributing feature through feature importance analysis, as shown in Figure 3. However, this approach did not yield significant enhancement.

Here Table 8 shows the performance (F1 score, Accuracy, Precision, and Recall) after using various algorithms and approaches suggested by

Table 7: Evaluation Metrics for Different Feature Combinations

Feature Combination	F1 Score	Accuracy	Precision	Recall
Angular Feature + Velocity Acceleration	0.43	0.57	0.60	0.39
Angular Feature + Velocity Acceleration + Trajectory Features	0.44	0.57	0.59	0.39
Angular Feature + Velocity Acceleration + Trajectory Features + Temporal Feature	0.44	0.57	0.59	0.39
Angular Feature + Velocity Acceleration + Trajectory Features + Energy Features	0.43	0.57	0.60	0.38
Angular Feature + Velocity Acceleration + Trajectory Features + Statistical Moments	0.44	0.57	0.59	0.39

LLMs such as ChatGPT, Gemini and Copilot. Here it is visible that LightGBM performs well in terms of the F1 Score and Accuracy of the model. Hence we finally used the LightGBM algorithm to propose our methodology. One of the most widely used visualization techniques for classification related performance is confusion matrix[16]. True Positive (TP), True Negative (TN), False Postive (FP), False Negative (FN) are the components of a confusion matrix.

Here, some of the used algorithms produced poor accuracy and F1 score. One of the possible reasons for the TAB Transformer that causes poor performance may be the inadequate data size. DT might have struggled to capture complex patterns in this data. On the other hand, SVM could not perform well because of its linear nature with non-linear decision boundaries in the data. It is clearly stated that Entity Embedding performs lowest in this context. One of the reasons could be over-fitting and lack of generalization to the used dataset.

6 Discussion

In this section, we discuss the advantages and disadvantages of using our proposed methodology using LLMs for feature engineering and algorithm

Table 8: Evaluation Metrics for Different Algorithms

Algorithms	**F1 Score**	**Accuracy**	**Precision**	**Recall**
LightGBM	0.46	0.58	0.62	0.41
TAB Transformer	0.31	0.50	0.30	0.28
FNN	0.31	0.50	-	-
CatBoost	0.29	0.51	0.65	0.29
Entity Embedding	0.15	0.18	0.17	0.19
DT	0.29	0.51	0.31	0.28
SVM	0.28	0.51	0.30	0.28
Random Forest Classification	0.29	0.51	0.31	0.29

selection in nursing activity recognition for ES.

6.1 Advantages of our Proposed Methodology

The main advantage of our proposed approach is its robustness in feature selection and algorithm optimization. By incorporating approaches suggested by LLMs such as ChatGPT, Gemini, and Copilot, we were able to identify the most impact features and best suitable algorithms, as evidenced by the enhanced F1 Score and accuracy from XGBoost to LightGBM. On the other hand, adaptability is also a major strength in our proposed methodology which incorporated new and effective feature sets such as Angular Features, Velocity Acceleration, and Trajectory Features. These features, suggested by LLMs helped our model to get higher accuracy in terms of activity recognition using skeleton-based video data.

6.2 Limitations and Disadvantages of our Proposed Methodology

Here one of the key limitations is the reliance on large labeled datasets. As the performance that is accuracy and F1 increase after incorporating new features, this may indicate that our model's performance may be contingent on the quality and availability of training data. Moreover, our model may influenced by biases if the LLMs such as ChatGPT, Gemini, and Copilot are trained on non-representative data. Furthermore, while using different LLMs, we experienced both similarities and different suggestions and approaches which created confusion and indecisive situations.

6.3 Comparative Analysis

In comparison to the traditional approach, our methodology shows promising results. However, our approaches using some of the advanced algorithms such as TAB Transformer, could not perform well in terms of accuracy and F1 Score. This may be due to the size of the dataset. On the other hand, LightGBM performed effectively in our approach, but its performance may vary as the size and diversity of the dataset increase and there is a potential that both accuracy and F1 Score might not maintain the same high standards with more varied datasets.

6.4 Leveraging LLMs for Enhanced Results

Leveraging suggestions from LLMs such as ChatGPT, Gemini, and Copilot significantly enhances the performances of our methodology. At first, before considering the addition of new features and algorithms, we tried to use approaches with existing features. Here, LLMs provided me with numerous suggestions, and Principal Component Analysis (PCA) was one of them which helped to slightly increase our model's performance. However, mainly by providing descriptions and structures of the dataset to LLMs, we got some features-related recommendations which are Angular Features, Velocity Acceleration, Trajectory Features, Temporal Features, Energy Features, Statistical Moments, Relative Positions, and Dynamic Time Warping (DTW). Considering some of the features provided us with higher accuracy and F1 Score. Also, LLMs help us with recommending various algorithms and a mix of algorithms which we implemented and analyzed the outputs according to Table VIII.

7 Conclusion

This study aimed to improve nursing activity recognition with the power of LLMs, a subset of GenAI which are now becoming more popular among researchers. Analyzing the potential research gap using LLMs to recognize activities of ES from skeleton data which are extracted from Video data. We have shown how LLMs such as ChatGPT, Gemini and Copilot helps to improve feature engineering, select the best-fit ML algorithm, and propose more structural programming solutions that enhance both accuracy and F1 Score. To automate nursing activity recognition in terms of ES, GenAI can be a good option to optimize the solution and gain more accuracy.

In this paper, we showed how two-way communication with LLMs can enhance the model performance in terms of F1 Score and Accuracy. Before using LLMs based suggestions and appraoches, we achieved an overall F1 Score of 0.31 with an Accuracy of 0.51. On the other hand, after using LLMs based suggessions and approaches we achieved a higher F1 Score of 0.58 with an Accuracy of 0.46. Therefore, it is clear that LLMs that is GenAI can improve our model performance by providing valuable guidance and insight like a human expert. While this study provides valuable insights for features selection and algorithms for skeleton-based datasets, it mainly focuses on identifying effective approaches rather than discussing the broader implications for future researchers. The analysis section offers a comprehensive comparison of different algorithms' performances on the dataset, aiding in understanding their efficacy.

The future version of this study aims to incorporate multi-modal data sources such as audio and vital body posture which may help to recognize nursing activities in ES more precisely. Moreover, we plan to explore the scalability of the GenAI frameworks. Finally, to enhance the accuracy and efficiency of nursing activity recognition our future aim will focus on integrating real-time feedback methods to aid and notify nurses during these critical ES procedures.

Appendix

Table 9: Experimental Setup Information

Parameter	**Value**
Language	Python
Device Configuration	Windows 10 (64-bit), 8GB RAM and Google Colab (Chrome Browser)
Libraries	Pandas, NumPy, Matplotlib, Scikit-learn, XGBoost, LightGBM, TensorFlow, Keras, PyTorch, CatBoost, TabNet, and TCN
Train-Test Split	5-fold cross-validation
Execution Time (First Part, XGBoost model)	CPU times: user 4 µs, sys: 0 ns, total: 4 µs Wall time: 7.15 µs
Execution Time (Second Part, LightGBM model)	CPU times: user 3 µs, sys: 0 ns, total: 3 µs Wall time: 6.91 µs

References

[1] K. Adachi, P. Lago, T. Okita, and S. Inoue. Improvement of human action recognition using 3d pose estimation. In M.A.R. Ahad, S. Inoue, D. Roggen, and K. Fujinami, editors, *Activity and Behavior Computing*, volume 204 of *Smart Innovation, Systems and Technologies*, chapter 2. Springer, Singapore, 2021.

[2] Amir Hossein Bayat, Mohammad Mahdi Arzani, Mahmood Fathy, Ali Matinnejad, Behrouz Minaei-Bidgoli, and Rahim Entezari. A probabilistic graphical model approach for human activity recognition using skeleton data. In *2016 2nd International Conference of Signal Processing and Intelligent Systems (ICSPIS)*, pages 1–5, Dec 2016.

[3] Xin Cao, Wataru Kudo, Chihiro Ito, Masaki Shuzo, and Eisaku Maeda. Activity recognition using st-gcn with 3d motion data. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, Ubi-

Comp/ISWC '19 Adjunct, page 689–692, New York, NY, USA, 2019. Association for Computing Machinery.

[4] Noriyo Colley, Shunsuke Komizunai, Shinji Ninomiya, Tadayoshi Asaka, Hiroki Mani, Hana Takizawa, Sarina Umesawa, Yuya Taira, Eri Murata, Misuzu Nakamura, Sozo Inoue, Satoshi Kanai, and Atsushi Konno. Hygienic manipulation evaluation system during endotracheal suctioning using bilateral dorsum manus total travel distance. In *2020 IEEE/SICE International Symposium on System Integration (SII)*, pages 500–503, Jan 2020.

[5] F. Faiz, Y. Ideno, H. Iwasaki, Y. Muroi, and S. Inoue. Multilabel classification of nursing activities in a realistic scenario. In M.A.R. Ahad, S. Inoue, D. Roggen, and K. Fujinami, editors, *Activity and Behavior Computing*, volume 204 of *Smart Innovation, Systems and Technologies*, chapter 17. Springer, Singapore, 2021.

[6] E. Ferrara. Genai against humanity: nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science*, 2024. Received: 10 November 2023; Accepted: 21 January 2024; Published: 22 February 2024.

[7] B. Friedrich, T.A.M. Orsot, and A. Hein. Using k-nearest neighbours feature selection for activity recognition. In M.A.R. Ahad, S. Inoue, D. Roggen, and K. Fujinami, editors, *Sensor- and Video-Based Activity and Behavior Computing*, volume 291 of *Smart Innovation, Systems and Technologies*, chapter 14. Springer, Singapore, 2022.

[8] Jiayuan Gao, Yingwei Zhang, Yiqiang Chen, Tengxiang Zhang, Boshi Tang, and Xiaoyu Wang. Unsupervised human activity recognition via large language models and iterative evolution. pages 91–95, April 2024.

[9] Plamen Hristov, Agata Manolova, and Ognian Boumbarov. Deep learning and svm-based method for human activity recognition with skeleton data. In *2020 28th National Conference with International Participation (TELECOM)*, pages 49–52, Oct 2020.

[10] Mofeoluwa Jide-Jegede and Tomiwa Omotesho. Harnessing generative ai for manufacturing innovation: Applications and opportunities. In *2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 568–572, 2024.

[11] Md. Eusha Kadir, Pritom Saha Akash, Sadia Sharmin, Amin Ahsan Ali, and Mohammad Shoyaib. Can a simple approach identify complex nurse care activity? In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, UbiComp/ISWC '19 Adjunct, page 736–740, New York, NY, USA, 2019. Association for Computing Machinery.

[12] Md. Ibrahim Mamun, Afroza Rahman, Md. Abdul Khaleque, Md. Abdul Hamid, and M. F. Mridha. Autilife: A healthcare monitoring system for autism center in 5g cellular network using machine learning approach. In *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*, volume 1, pages 1501–1506, July 2019.

[13] Hoang Anh Vy Ngo, Haru Kaneko, Iqbal Hassan, Elsen Ronando, Milyun Ni'ma Shoumi, Ryuya Munemoto, Tahera Hossain, and Sozo Inoue. Summary of the nurse care activity recognition challenge using skelton data from video with generative ai. *International Journal of Activity and Behavior Computing*, 2024, 2024.

[14] Hoang Anh Vy Ngo, Quynh N Phuong Vu, Noriyo Colley, Shinji Ninomiya, Satoshi Kanai, Shunsuke Komizunai, Atsushi Konno, Misuzu Nakamura, and Sozo Inoue. Toward recognizing nursing activity in endotracheal suctioning using video-based pose estimation. *International Journal of Activity and Behavior Computing*, 2024(1):1, 2024.

[15] Petar Nikolov, Ognian Boumbarov, Agata Manolova, Krasimir Tonchev, and Vladimir Poulkov. Skeleton-based human activity recognition by spatio-temporal representation and convolutional neural networks with application to cyber physical systems with human in the loop. In *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, pages 1–5, July 2018.

[16] Luc-Etienne Pommé, Romain Bourqui, Romain Giot, and David Auber. Relative confusion matrix: Efficient comparison of decision models. In *2022 26th International Conference Information Visualisation (IV)*, pages 98–103, July 2022.

[17] S. Reddy. Generative ai in healthcare: an implementation science informed translational path on application, integration and governance. *Implementation Science*, 19:27, 2024. Received: 17 August 2023; Accepted: 06 March 2024; Published: 15 March 2024.

[18] A. Sarkar, S.K.S. Hossain, and R. Sarkar. Human activity recognition from sensor data using spatial attention-aided cnn with genetic algorithm. *Neural Computing and Applications*, 35:5165–5191, Mar 2023. Received: 08 March 2022; Accepted: 29 September 2022; Published: 26 October 2022.

[19] Mohsin Raza Siyal, Mansoor Ebrahim, Syed Hasan Adil, and Kamran Raza. Human action recognition using convlstm with gan and transfer learning. In *2020 International Conference on Computational Intelligence (ICCI)*, pages 311–316, 2020.

[20] Md. Hasan Tarek, Md. Eusha Kadir, Mahir Mahbub, Pritom Saha Akash, Amin Ahsan Ali, and Mohammad Shoyaib. Mutual information based feature selection for nurse care activity recognition. In *2020 Joint 9th International Conference on Informatics, Electronics Vision (ICIEV) and 2020 4th International Conference on Imaging, Vision Pattern Recognition (icIVPR)*, pages 1–6, Aug 2020.

[21] Eteka Sultana Tumpa and Krishno Dey. A review on applications of machine learning in healthcare. In *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, pages 1388–1392, April 2022.

[22] Ping Yu, Hua Xu, Xia Hu, and Chao Deng. Leveraging generative ai and large language models: A comprehensive roadmap for healthcare integration. *Healthcare*, 11(20), 2023.