Optimizing Endotracheal Suctioning Classification: Leveraging Prompt Engineering in Machine Learning for Feature Selection

Mahera Roksana Islam
Department of Electrical and Electronic Engineering
University of Dhaka
Dhaka, Bangladesh
maherarislam@gmail.com

Anik Mahmud Ferdous
Department of Electrical and Electronic Engineering
University of Dhaka
Dhaka, Bangladesh
anikmahmud0001@gmail.com

Shahera Hossain
Department of Computer Science and Engineering
University of Asia Pacific
Dhaka, Bangladesh
shaherahossain@gmail.com

Md Atiqur Rahman Ahad
Department of Computer Science and Digital Technologies
University of East London
London, United Kingdom
mahad@uel.ac.uk

Fady Alnajjar*
Department of Computer Science and Software Engineering
College of Information Technology, UAE University
Al Ain, UAE
fady.alnajjar@uaeu.ac.ae

***Abstract*—In a world with an increasingly aging population, there is a critical demand for more skilled professionals in the nursing industry. AI-based systems can be more effective than traditional methods, which rely on in-person assistance, by accurately identifying nursing activities and assessing nursing trainees. Such systems can help train them to become proficient in their roles.This paper addresses classifying activities in one such system, endotracheal suctioning, using skeletal keypoint data of the subject performing the procedure. A multi-step structured prompt engineering method was established and utilized on several LLMs to select or calculate critical features from the data. The features were subsequently passed onto several tuned machine-learning models to obtain the final results. A tuned XGBoost prevailed across all models, achieving 90% accuracy on the validation set.**

***Index Terms*—Human Activity Recognition, Large Language Model, Generative AI, Machine learning, Nurse-care**

*

I. Introduction

In the context of nursing education, accurate recognition of training activities is essential for effective skill assessment and feedback. Traditional methods for activity recognition often rely on manual observation and annotation, which can be time-consuming, subjective, and prone to errors [1]. To overcome such limitations and to reduce the need for manual intervention, accurate AI-based human activity recognition systems are needed. However, the complexity of certain nursing activities, like endotracheal suctioning (ES), which involves intricate body movements and interactions with medical equipment, poses formidable challenges to accurate action recognition [2]. Furthermore, occlusion due to limitations of camera frames and positions and the presence of background non-primary subjects compounds this problem [3].

Recent advancements in artificial intelligence (AI) and computer vision technologies have revolutionized the classification of human motion data. The use of skeletal keypoints in a time frame to classify actions, aided by the rise of generative AI, particularly in large language models, has shown promising results. These technologies can be prompted to perform complex data analysis and detect underlying trends and features, marking a significant leap in the field of AI-based activity recognition.

Endotracheal suctioning (ES) is a crucial intervention in patients requiring mechanical ventilation to clear airway secretions and maintain optimal respiratory function. This multi-step procedure involves the insertion of a suction catheter into the endotracheal tube to remove mucus and debris, reducing the risk of airway obstruction and improving gas exchange. Proper technique, timing, and monitoring are essential to minimize complications and ensure patient safety during endotracheal suctioning procedures [8]. So far, no significant research or system pipelines have been proposed for the use of AI in automating the assessment of the procedure.

Recent advancements in natural language processing, particularly in the realm of large language models (LLMs), have allowed the usage of natural language or ”prompts” to

[*] Corresponding: fady.alnajjar@uaeu.ac.ae

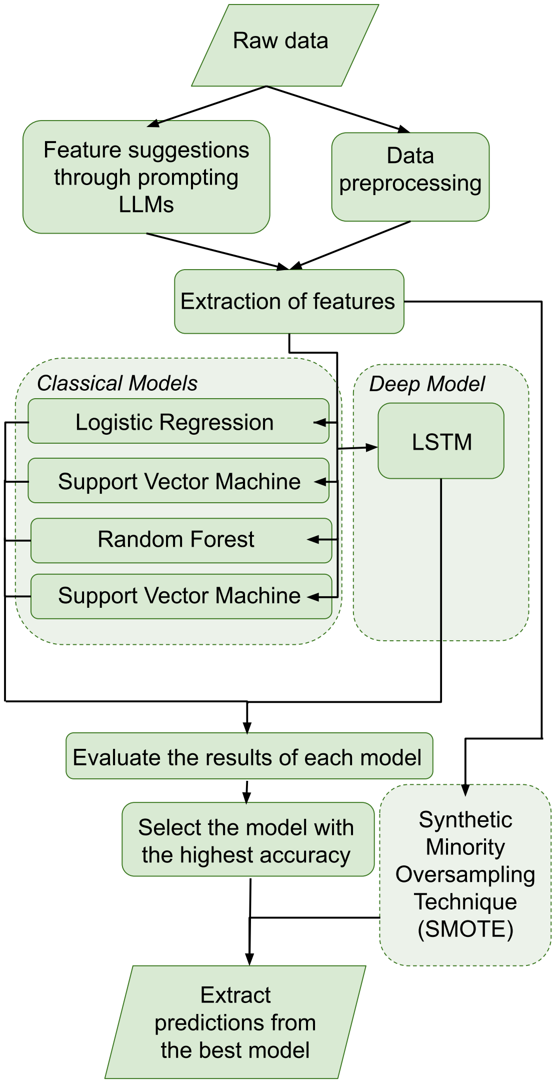

Raw data
Feature suggestions through prompting LLMs
Data preprocessing
Extraction of features
Classical Models
Logistic Regression
Support Vector Machine
Random Forest
Support Vector Machine
Deep Model
LSTM
Evaluate the results of each model
Select the model with the highest accuracy
Synthetic Minority Oversampling Technique (SMOTE)
Extract predictions from the best model

Fig. 1. Flowchart of the proposed methodology.

tune these models to accomplish specific tasks [5]. In this paper, conversational models like Claude 2.0 and GPT 3.5 are prompted in a step-by-step logically structured approach called Chain-of-thought strategy to obtain suggestions for feature extraction based on the skeletal data [12]. In this paper, we propose an approach for activity recognition of nurse training activities in ES procedure using skeletal data as illustrated in Fig. 1. Specifically, we focus on the following key contributions:

1) Prompting LLMs to suggest a subset of relevant features based on our raw data.
2) Developing a machine learning model to accurately classify nurse-care activities.
3) Ameliorating the obstacles posed by imbalanced datasets by oversampling minority classes.

The remainder of this paper is organized as follows: Section II presents related work, Section III shows dataset information, the classification processes used on the activity, and the data pre-processing information. Section IV discusses the methods utilized in the classification task. Section V presents the results obtained from our approach.

II. Related Works

Endotracheal suctioning is a procedure involving the removal of pulmonary secretions, which is necessary to ensure sufficient oxygen reaches the patients lungs, specially in intensive care units. To ensure this procedure takes place smoothly, nurses need to be trained well to ensure they are aware of correct techniques [11]. This can be ensured by the help of proper activity recognition systems, allowing nurses to learn such practices long before it is ever applied in life threatening situations. Endotracheal suctioning, though uncomfortable, is deemed necessary by patients to ease breathing [10].

On the activity recognition side, much progress has been made. Specifically for using skeletal data for nursecare activity recognition, multimodal transformer based network can be used to bring out features from skeletal joints and acceleration data to perform nurse care activity identification [2]. Besides this, deep learning models have been found to be better adapted to the field of multimodal datasets compared to hand crafted features [6].

Joint position data can be utilized for action recognition. Previous studies have explored using statistical features like mean, minimum, and maximum joint positions as well as applying principal component analysis (PCA) to the joint position data to extract relevant features for action classification [21].

Previous approaches to action recognition have also utilized Fast Fourier Transform (FFT) on the time series of joint positions. Specifically, the first five FFT components extracted from the joint position sequences are used as a feature vector, which is then passed into a neural network for action classification. However, this did not perform well in recognizing complex actions involving different body parts [22].

In the realm of skeletal data in nursing, most existing skeletal data models operate in one of two ways: (1) They utilize manually calculated features, which are then inputted into an ensemble of decision trees [23] or neural networks [24], or (2) they directly feed the data into sequential models [25]. A few models use convolutional neural networks (CNNs) to extract features from raw data [26]. However, none of such models extract features in relation to the characteristic nature of the data and the relevance of its contents. This paper is distinctive as we use LLMs to obtain prompted feature suggestions which is prompted to take into account both the peculiarities of the dataset and the context of the information it represents.

III. Dataset

A. Data Collection

The dataset used for this paper consists of ten nurses and twelve nursing students performing the procedure for endotracheal suctioning (ES) on the simulation system, ESTE-SIM [19]. The camera was positioned directly in front of the nurse performing the procedure on a mannequin laid on a hospital bed. Due to the lack of multiple camera positioning, only one perspective of the activity performed could be captured, and often led to occlusion due to the nurse often moving out of the camera frame.

Each of the twenty-four subjects involved performed the ES procedure twice as follows: tracheal suctioning, positioning, lung auscultation, tracheal suctioning, and lung auscultation - resulting in 44 videos.

Written consents were collected and the ethical approval of the data collection was obtained from Hokkaido University, Faculty of Health Sciences. The video camera used in the experiment was SONY HANDYCAM HDR-PJ680 with the focal length of 1.9-57.0mm. The frame per second of video is 30 and the image size is 1920×1080 [20].

B. Data Description

Each activity performed for this dataset is labeled by numbers in Table I. The ”Others” tag represents activity that does not fall under any of the designated 8 activities. It can be seen from Fig. 2 that the dataset is quite imbalanced with Activity 4 (”Catheter Disinfection”) occurring the most frequently and Activity 6 (”Positioning”) occurring the least. From Fig. 3, it can be deduced that the time required for each activity differs widely with Activity 0 (”Catheter Preparation”) taking the most time in total.

TABLE I
ACTIVITIES OF ENDOTRACHEAL SUCTIONING (ES) PROCEDURES AND THEIR RESPECTIVE LABELS

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

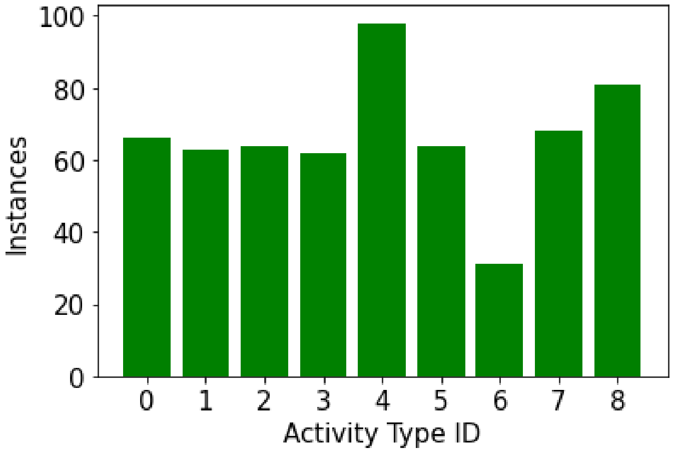


Fig. 2. Number of Instances of Activity Labels

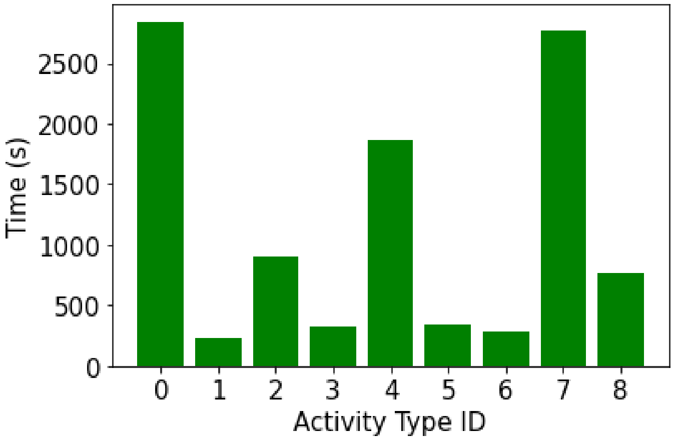


Fig. 3. Total Time of Activity Labels

C. Data Splitting and Processing

All of the videos were passed through YOLOv7 to extract the keypoint data of the video subjects. From each frame of each video, 17 keypoint sets of x and y coordinates of the main subject and their respective confidence scores were obtained. Data was processed to remove any background subjects. The keypoints are as follows: nose, pairs of eye, ear, shoulder, elbow, wrist, hip, knee, ankle. Among the 44 videos obtained, 12 videos (∼27%) were kept for testing and 6 videos (∼13.6%) were kept for validation. The rest of the 26 videos were used for training the models.

IV. METHOD

The following section details our approach to prompting LLMs for feature extraction, and using those features to develop a machine learning model for activity classification.

A. Preprocessing

For the purposes of simplification, speed, and limited data storage capability, the keypoint data for body parts below the torso were removed during preprocessing. As the videos portrayed the main subject's body occluded by a mannequin from the waist down, the keypoint data for coordinates of knees and ankles were deemed unnecessary. Similarly, the confidence scores were also removed. To lessen the noise and random fluctuations in the data and to make the apparent data trends more visible, the data was smoothed by taking the mean of every 3 seconds. Fig. 4 shows an example of the x and y coordinate values of the left wrist of a subject before and after smoothing was performed.

B. Feature Selection Using Prompt Engineering

LLMs were used throughout the course of this approach, for generating subsets of potential features. To ensure control across the LLMs, they were guided through a strictly methodical process step-by-step to infer and suggest the most relevant features from the keypoints data. Initially, the LLMs were prompted to suggest general features that may be relevant

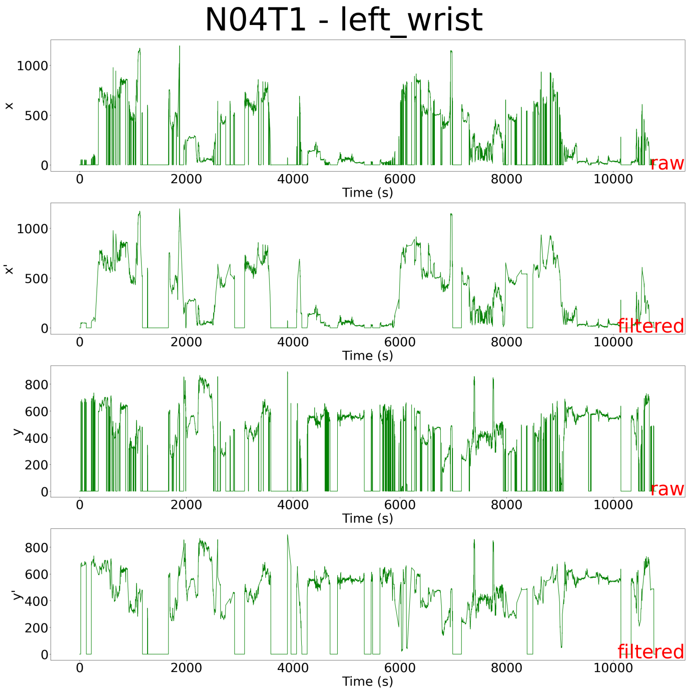


Fig. 4. Data of left wrist keypoints of a sample subject before and after smoothing.

to human activity recognition. After that, a 4-part structured process to prompt the details of features to be used in the model:

- Introduction, I
- Context, $C(x_i)$
- Chain of Thought (CoT), T
- Question, Q

In this prompting process, the LLMs were first prompted with the introduction, I, of the problem which was to analyze the keypoint descriptors and activity label annotations, and select features to distinguish each activity. Then the keypoint data and annotations were fed as context, $C(x_i)$ where x_i was a data point. We, then, prompted the LLMs to analyze and determine which movements in particular (e.g. which joint angles or distances) and its relevant parameters are important for classification, and to determine their respective roles under each activity label. This was a multi-step process and methods from Chain-of-Thought (CoT) prompting were utilized [12]. Lastly, we directed the LLMs to choose the activity labels, from 0 to 7, which matches the activity most closely. The ones that could not be classified were marked as Others (label 8).

The described prompt is represented as shown in Equation 1.

$$\mathbf{E}(\mathbf{x}_i) = \{\mathbf{I}, \mathbf{C}(\mathbf{x}_i), \mathbf{T}, \mathbf{Q}\} \tag{1}$$

The feature set, $\mathcal{F}$, is obtained from passing the prompts through LLMs, where each structured output is as shown in Equation 2.

$$\mathbf{f}_i = \mathrm{LLM}(\mathbf{E}(\mathbf{x_i})) \tag{2}$$

Name of the Feature		Description of Importance	Suggested by	
			GPT 3.5	Claude 2.0
Basic statistical features	Mean	Arithmetic average of data in each frame	☑	☑
	Variance	Measure of spread of a feature of a frame	□	☑
	Standard deviation	Squared root of variance	☑	□
	Maximum value	Maximum value of a frame	□	☑
	Minimum value	Minimum value of a frame	□	☑
	Median	Median value of a feature of a frame	☑	☑
	Sum	Sum of all values of a feature of a frame	□	☑
Joint angles	Between elbow, shoulder and hip (for both left and right side)	Change in angles of certain joints across frames provides spatial understanding of an activity and helps to differnetiate one activity from another.	☑	□
	Between wrist, elbow and shoulder (for both left and right side)		☑	☑
	Between left elbow, left, and right shoulder		□	☑
	Between right elbow, right, and left shoulder		□	☑
	Between right shoulder, nose, left shoulder		□	☑
	Between hip center, shoulder, elbow (for both left and right side)		☑	□
Joint distances	From shoulder to wrist (for both left and right side)	Similar to joint angles, particular joint distances provide information about particular momental characteristics of each acitvity and helps to differentiate between multiple acitvities.	□	☑
	From hip center to elbow (for both left and right side)		☑	☑
	From hip center to wrist (for both left and right side)		☑	□
	From right wrist to left wrist		□	☑
	From right hip to left hip		□	☑
	From hip side to wrist (for both left and right side		□	☑
Velocity		Calculated as the rate of change of movements in a frame	☑	☑
Acceleration		Calculated as the rate of change of velocity in a frame	☑	☑
Jerk		Calculated as the rate of change of acceleration in a frame	☑	□

Fig. 5. Features obtained from prompting LLMs.

The features obtained can be grouped into following categories:

- **Basic statistical features:** For each frame, the following are calculated: mean (arithmetic), variance, standard deviation, maximum value, minimum value, median, sum.
- **Velocity, acceleration, and jerk:** These help to understand the rate of change of movements across several time frames.
- **Joint angles:** Angle between two joints, a and b, were calculated using the equation for dot product as represented in Equation 3.

  $$\theta = \arccos\left(\frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}||\mathbf{b}|}\right) \tag{3}$$

  In total, **nine** angles were obtained from prompting LLMs as illustrated on Fig. 5.
- **Joint distances:** The distances are calculated using Euclidean formula for distances, x and y as represented in Equation 4.

  $$d = \sqrt{(x)^2 + (y)^2} \tag{4}$$

  In total, **ten** distances were obtained from prompting LLMs as illustrated on Fig. 5.

Joint angles and distances in combination builds spatial representation of the skeletal data and help to understand the movements and changes in joints and body parts with each distinctive activity [6]. As joint distances differ significantly amongst individuals, they had to be normalized using spine as the reference [13]. Due to the absence of keypoint data, the spinal distance was estimated from the center of the shoulders to the center of hips, which was calculated by taking the midpoint of the left and right hip point data. Overall 861 features were extracted for each frame.

C. Model Tuning

To classify actions based on the generated features, four classical models and one deep learning model were tuned and tested on the dataset.

1) Classical Models: By prompting the LLM's repeatedly, we attained sets of possible hyperparameter values for each classical model. Then through searching through the possible combinations of hyperparamter values (using grid search and randomized search), we obtained most suitable values for each model. The models, hyperparameter options and their tuned hyperparameters are shown in Table II.

2) Deep Learning Model: A simple Long Short Term Memory (LSTM) model was run on the dataset. The LSTM consisted of one hidden layer, followed by a Dropout layer, followed by a Dense layer with ReLu activation function [14], before being passed off to the output. Each of the keypoint skeletal part, arranged sequentially, were fed into the model with a window size of 3 seconds.

D. Testing Feature Subsets

Each subset of features found from each LLM was tested against the tuned model which returned the best result. The entire set of features and no features were also run on the model to analyse the differences in impact on results based on the features suggested by LLMs.

TABLE II
CLASSICAL MODELS USED WITH THEIR HYPERPARAMETERS SET

Model	Parameters Tuned	Tuning Selections	Tuned Value
Random Forest	n_estimators	[10, 50, 100, 250]	100
	max_depth	[5, 10, 20]	20
Classifier Support Vector Machine	C parameter	[10**-2, 10**-1, 10**0, 10**1, 10**2]	10**2
	class_weight	[None, 0:1,1:5, 0:1,1:10, 0:1,1:25]	0:1,1:5
Logistic Regression	C parameter	[10**-2, 10**-1, 10**0, 10**1, 10**2]	10**1
	class_weight	[None, 0:1,1:5, 0:1,1:10, 0:1,1:25]	None
XGBoost	max_depth	[5, 10, 15]	5
	min_child_weight	[2, 5, 10, 30]	2
	subsample	[0.8, 0.5]	0.5
	colsample_bytree	[0.7, 0.5]	0.7
	learning_rate	[0.1, 0.01, 0.001]	0.1
	n_estimators	[100, 350, 500, 900]	900

E. Handling of Class Imbalance Data

As seen from Fig. 2, the instances of each activity labels vary greatly. Activity 4 ('Catheter Disinfection') and 8 ('Others') occur the most whereas activity 6 ('Positioning') occurs the least. Due to such imbalance, the model has less data of the lower-occurring classes to learn from which might lead to misidentification and overall lower accuracy. So, Synthetic Minority Over-sampling Technique (SMOTE) [15] from the imblearn package in Python was used to generate synthetic data points, and rectify the problem of large class imbalance.

F. Performance Evaluation

The performance of each model is primarily evaluated with respect to their F1-score. Precision is the ratio of true positives and all positive values [16]. Recall is the ratio of true positives and all relevant values (true positives and false negatives). F1-score is calculated by taking the harmonic average of precision and recall. Macro average is calculated by finding the mean of all F1-scores across activity labels. Accuracy (or micro average) is found by finding the global F1-score across all classes. Weighted average is found by averaging the activity labels with support instances as weights of each activity label.

V. RESULTS AND DISCUSSION

A. Classical Models

From the results on the validation set, as illustrated in Table III, it can be seen that the XGBoost performed best with an accuracy of 0.89, while logistic regression performed the worst. The baseline Random Forest model provides an accuracy of 0.64 and is eventually tuned to 0.85. It can be inferred that tree-based models like Random Forest and XGBoost performed much better than the other models. This may have been because the principal data source is tabular, and tree-based methods are well-known for being effective on structured data due to their superior ability to handle non-linear relationships and their robustness to irrelevant features and variables [17].

TABLE III
CLASSICAL MODEL PERFORMANCE.

Model	Accuracy	Macro Average	Weighted Average
Logistic Regression	0.73	0.63	0.72
Random Forest (Tuned)	0.85	0.75	0.83
Random Forest (Baseline)	0.64	0.59	0.63
XGBoost	**0.89**	**0.80**	**0.89**
Support Vector Machine	0.73	0.64	0.73

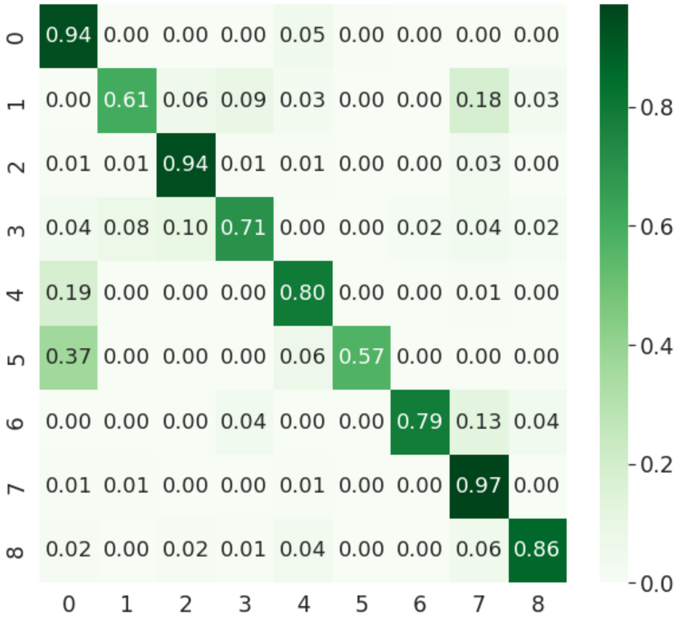


Fig. 6. Confusion Matrix of tuned XGBoost Model.

1) Results from Feature Subsets: When each feature subsets obtained from each LLM were passed onto the model with the highest accuracy, it resulted in Table IV. Combining feature subsets obtained from both LLMs results in the highest accuracy of 0.89. This is a significant improvement compared to results (62%) where no LLMs were utilized to extract features. It is also noticeable that the feature subset obtained from GPT 3.5 achieved an accuracy of 0.75 with 342 features, while Claude 2.0 achieved a marginally higher result of 0.77 with more than twice the number of features.

2) Results with SMOTE: From the confusion matrix of XGBoost in Fig. 6, it can be seen that the model performed the worst on labels with the lowest support instances, namely labels 1 and 5. This is because the model has fewer examples to train and learn from to provide optimal classification. To rectify such imbalance of labels, SMOTE was used, resulting in the outputs in Fig. 7.

TABLE IV
RESULTS OF FEATURE SUBSETS PASSED THROUGH XGBOOST.

Feature extraction method	Number of features extracted	Accuracy
No LLMs	-	0.62
GPT 3.5	342	0.75
Claude 2.0	702	0.77
Both GPT 3.5 and Claude 2.0	**861**	**0.89**

TABLE V
COMPARISON BETWEEN XGBOOST MODEL WITH AND WITHOUT SMOTE DATA.

Activity Label	XGBoost		SMOTE with XGBoost	
	F1 score	Support	F1 score	Support
0	0.90	526	0.90	555
1	0.65	33	0.76	34
2	0.94	148	0.94	168
3	0.76	51	0.72	52
4	0.82	285	0.89	352
5	0.70	49	0.82	56
6	0.86	47	0.81	50
7	0.96	501	0.94	540
8	0.90	109	0.87	136
accuracy	0.89	1749	**0.90**	1943
macro average	0.80	1749	**0.85**	1943
weighted average	0.89	1749	**0.90**	1943

It is noticed from Table V that the accuracy with SMOTE marginally increased, especially for low-instance labels due to a larger number of support instances due to synthetic generation, thus increasing the macro average from 80% to 85%.

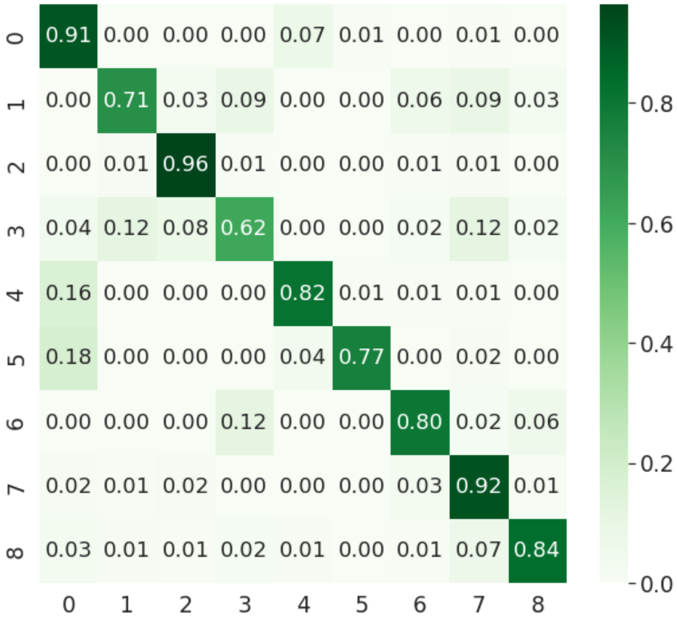
Overall, it is noticed that the accuracy and weighted averages are more than the macro averages across all models and sampling techniques. This is because the macro average calculates each class independently and averages to give the final score. It treats each class equally, regardless of its support. In this imbalanced dataset, since the minority classes have poor performance (low precision, recall, etc.), they drag down the macro average [18].

B. Deep Learning Model

The LSTM model was run 10 times, and resulted in an average accuracy of 65.12%, which was not nearly comparable to the accuracy we obtained from decision trees. The accuracy can be increased if the architecture is finetuned to include more layers, it's learning rate, sequence length, batch size, and hyperparameters are optimized.

VI. CONCLUSION

Due to the imperative need for automated systems in nurse-care services, leveraging machine learning models becomes essential for enhancing efficiency, accuracy, and patient outcomes. In this paper, we propose a machine learning model



	0	1	2	3	4	5	6	7	8
0	0.91	0.00	0.00	0.00	0.07	0.01	0.00	0.01	0.00
1	0.00	0.71	0.03	0.09	0.00	0.00	0.06	0.09	0.03
2	0.00	0.01	0.96	0.01	0.00	0.00	0.01	0.01	0.00
3	0.04	0.12	0.08	0.62	0.00	0.00	0.02	0.12	0.02
4	0.16	0.00	0.00	0.00	0.82	0.01	0.01	0.01	0.00
5	0.18	0.00	0.00	0.00	0.04	0.77	0.00	0.02	0.00
6	0.00	0.00	0.00	0.12	0.00	0.00	0.80	0.02	0.06
7	0.02	0.01	0.02	0.00	0.00	0.00	0.03	0.92	0.01
8	0.03	0.01	0.01	0.02	0.01	0.00	0.01	0.07	0.84

Fig. 7. Confusion Matrix of tuned XGBoost Model on SMOTE data.

with features selected through prompt engineering conversational large language models (LLMs) to classify the activities in endotracheal suctioning (ES) procedure with 90% accuracy.

Our research was limited due to the low number of subjects in our dataset, resulting in a smaller sample size for training. Moreover, since YOLOv7 was used to extract skeletal key points of human subjects from videos, the data was limited to coordinates of 17 points, with no points for hip center and spine - parts that we had to interpolate by taking an average of adjacent points. In the future, models with more coordinate points can be used for more accurate predictions.

Since the dataset is heavily imbalanced, SMOTE was used, which resulted in marginally improved results. However, SMOTE generated synthetic examples based solely on the existing minority class instances, potentially losing valuable information in the original data. This could have resulted in the creation of primarily noisy or irrelevant synthetic samples. Furthermore, it may not have been as effective as the class imbalance, which may have been caused by inherent differences in the class distributions (i.e., a particular activity may naturally occur seldomly).

Although we used four classifiers and a sequential model for classification, the challenge can be extended across other classifiers and neural networks. This study can also be extended across different and more updated conversational LLMs like GPT 4.0 [27]. For future works, advanced models like transformers can be used to sequentially model different joints and use attention networks to weigh the importance of different parts of the input. Furthermore, generative adversarial networks (GANs) can be implemented to overcome occlusion posed by image subjects being out of frame. Such models can further assist in reconstructing the subject's entire skeletal morphology, which can lead to better feature extraction and model prediction.

References

[1] Faiz, F., Ideno, Y., Iwasaki, H., Muroi, Y., Inoue, S. (2021). Multilabel Classification of Nursing Activities in a Realistic Scenario. In: Ahad, M.A.R., Inoue, S., Roggen, D., Fujinami, K. (eds) Activity and Behavior Computing. Smart Innovation, Systems and Technologies, vol 204. Springer, Singapore. https://doi.org/10.1007/978-981-15-8944-7_17

[2] Momal Ijaz, Renato Diaz, Chen Chen: Multimodal transformer for nursing activity recognition (2022). DOI https://doi.org/10.48550/arXiv.2204.04564

[3] "Human activity recognition challenge," SpringerLink, https://link.springer.com/book/10.1007/978-981-15-8269-1? (accessed Apr 30, 2024).

[4] J. Guo, V. Mohanty, H. Hao, L. Gou, and L. Ren, "Can LLMS infer domain knowledge from code exemplars? A preliminary study," Companion Proceedings of the 29th International Conference on Intelligent User Interfaces, Mar. 2024. doi:10.1145/3640544.3645228

[5] Language models are unsupervised multitask learners - 2018, https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf (accessed Apr. 29, 2024).

[6] S. Sarker, S. Rahman, T. Hossain, Syeda Faiza Ahmed, L. Jamal, and Atiqur Rahman Ahad, "Skeleton-Based Activity Recognition: Preprocessing and Approaches," Intelligent systems reference library (Print), pp. 43–81, Jan. 2021, doi: https://doi.org/10.1007/978-3-030-68590-4_2.

[7] T. Hossain, S. Sarker, S. Rahman, and Atiqur Rahman Ahad, "Skeleton-Based Human Action Recognition on Large-Scale Datasets," Intelligent systems reference library (Print), pp. 125–146, Jan. 2021, doi: https://doi.org/10.1007/978-3-030-75490-7_5.

[8] C. J. Wood, "Endotracheal suctioning: a literature review," Intensive and Critical Care Nursing, vol. 14, no. 3, pp. 124–136, Jun. 1998, doi: https://doi.org/10.1016/s0964-3397(98)80375-3.

[9] Chen, Weiming; Jiang, Zijie; Guo, Hailin; Ni, Xiaoyang (2020). Fall Detection Based on Key Points of Human-Skeleton Using OpenPose. Symmetry, 12(5), 744–. doi:10.3390/sym12050744

[10] Pedersen, C.M., Rosendahl-Nielsen, M., Hjermind, J., Egerod, I.: Endotracheal suctioning of the adult intubated patient—what is the evidence? Intensive and Critical Care Nursing 25(1), 21–30 (2009). DOI https://doi.org/10.1016/j.iccn.2008.05.004. URL https://www.sciencedirect.com/science/article/pii/S0964339708000566

[11] Day, T., Farnell, S., Wilson-Barnett, J.: Suctioning: a review of current re- search recommendations. Intensive and Critical Care Nursing 18(2), 79– 89 (2002). DOI https://doi.org/10.1016/S0964-3397(02)00004-6. URL https://www.sciencedirect.com/science/article/pii/S0964339702000046

[12] Jason Zhanshun Wei et al., "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," Jan. 2022, doi: https://doi.org/10.48550/arxiv.2201.11903.

[13] R. Li, H. Fu, W. Lo, Z. Chi, Z. Song, and D. Wen. Skeleton-based action recognition with key-segment descriptor and temporal step matrix model. IEEE Access, 7:169782–169795, 2019.

[14] K. Fukushima, "Visual Feature Extraction by a Multilayered Network of Analog Threshold Elements," IEEE Transactions on Systems Science and Cybernetics, vol. 5, no. 4, pp. 322–333, 1969, doi: https://doi.org/10.1109/tssc.1969.300225.

[15] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, vol. 16, no. 16, pp. 321–357, Jun. 2002, doi: https://doi.org/10.1613/jair.953.

[16] David, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," Oct. 2020, doi: https://doi.org/10.48550/arxiv.2010.16061.

[17] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on tabular data?," arXiv:2207.08815 [cs, stat], Jul. 2022, Available: https://arxiv.org/abs/2207.08815

[18] T. Gowda, W. You, C. Lignos, and J. May, "Macro-Average: Rare Types Are Important Too." Available: https://arxiv.org/pdf/2104.05700.pdf

[19] H. A. V. Ngo et al., "Summary of the Nurse Care Activity Recognition Challenge Using Skelton Data from Video with Generative AI," International Journal of Activity and Behavior Computing, vol. 2024, 2024.

[20] H. A. V. Ngo et al., “Toward Recognizing Nursing Activity in Endotracheal Suctioning Using Video-based Pose Estimation,” International Journal of Activity and Behavior Computing, vol. 2024, no. 1, pp. 1–20, 2024. doi:10.60401/ijabc.1

[21] V. R. Reddy, T. Chattopadhyay, ”Human activity recognition from kinect captured data using stick model,” In Human-Computer Interaction. Advanced Interaction Modalities and Techniques, Springer, pp. 305–315 , 2014.

[22] Mario Martínez-Zarzuela and Francisco J Díaz-Pernas and A. Tejeros-de-Pablos and David González-Ortega and Míriam Antón-Rodríguez, ”Action recognition system based on human body tracking with depth images,” Advances in Computer Science: an International Journal, vol. 3(1), pp. 115–123, 2014. URL: https://api.semanticscholar.org/CorpusID:6122142

[23] D. A. Adama, A. Lotfi, C. Langensiepen, K. Lee, and P. Trindade, ”Human activity learning for assistive robotics using a classifier ensemble,” Soft Computing, vol. 22, no. 21, pp. 7027–7039, Jul. 2018. doi:10.1007/s00500-018-3364-x

[24] E. Mathe, A. Maniatis, E. Spyrou, and P. Mylonas, ”A deep learning approach for human action recognition using skeletal information,” Advances in Experimental Medicine and Biology, pp. 105–114, 2020. doi:10.1007/978-3-030-32622-7_9

[25] R. Cui, A. Zhu, G. Hua, H. Yin, and H. Liu, ”Multisource learning for skeleton-based action recognition using Deep LSTM and CNN,” Journal of Electronic Imaging, vol. 27, no. 04, p. 1, Aug. 2018. doi:10.1117/1.jei.27.4.043050

[26] S. K. Yadav, K. Tiwari, H. M. Pandey, and S. A. Akbar, ”Skeleton-based human activity recognition using CONVLSTM and guided feature learning,” Soft Computing, vol. 26, no. 2, pp. 877–890, Oct. 2021. doi:10.1007/s00500-021-06238-7

[27] Z. Zhao, O. Mubin, F. Alnajjar, and L. Ali. ”A pilot study of measuring emotional response and perception of LLM-generated questionnaire and human-generated questionnaires.” Scientific Reports 14, no. 1 (2024): 2781.