2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) | 979-8-3315-8618-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/EMBC58623.2025.11253978

# Surgical Workflow Analysis: An Explainable Approach

Christos Spiliadis[1], Yiheng Chang[2], Justin Dauwels[3], Chavdar Bachvarov[4], John J. van den Dobbelsteen[5], Benno H. W. Hendriks[6], Maarten Van der Elst[7], and Markku Eskola[8]

Abstract— Surgical workflow analysis optimizes efficiency, resource use, and patient safety in catheterization labs. Traditional manual methods are labour-intensive and inconsistent, driving the need for automated solutions that utilize machine learning and computer vision. This thesis introduces an explainable two-stage model for workflow analysis using ceiling-mounted cameras. The approach combines a YOLOv8 object detection model with a Gaussian Mixture Model - Hidden Markov Model (GMM-HMM). The first stage detects key objects for input into the second stage, where the GMM-HMM infers workflow phases by modelling spatial and temporal dynamics for real-time classification. Validation on two hospital datasets achieves 95.2% accuracy for the RdGG dataset and 95.4% for HH Tampere, demonstrating generalizability across environments. Experimental results show high accuracy in detecting workflow phases, highlighting explainability and robustness. The combined efficiencies of YOLOv8 and GMM-HMM allow for precise phase transition identification. The model's real-time application and adaptability across hospitals suggest its clinical implementation potential. This research furthers automated workflow analysis by enhancing interpretability and adaptability. Future work aims to improve robustness against occlusions, integrate audio data, and explore applications in other surgical settings.

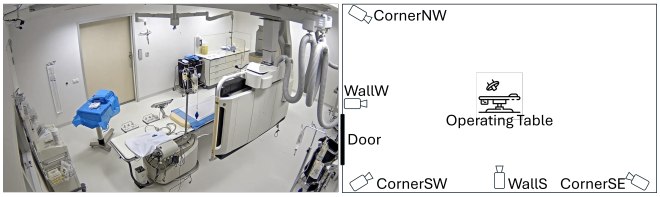


Fig. 1. Left: View from the camera at the South East corner in RdGG. Right: Layout of RdGG Cath Lab.

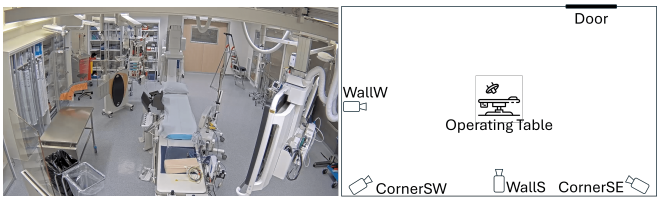


Fig. 2. Left: View from the camera at the South wall in HH Tampere. Right: Layout of HH Tampere Cath Lab.

## I. INTRODUCTION

Cardiovascular diseases are among the leading causes of death globally and a significant contributor to hospitalizations [1]. To address this, healthcare systems have increasingly turned to minimally invasive procedures performed in Catheterization Laboratories (Cath Labs) [2]. These specialized hospital units offer advanced imaging systems and tools that allow physicians to diagnose and treat heart-related conditions in a controlled environment. Procedures such as coronary angiography and percutaneous coronary interventions (PCI), which are among the most common procedures performed in Cath Labs, are carried out by multidisciplinary teams, leveraging state-of-the-art C-Arm angiographers for real-time imaging and precise catheter guidance.

The growing demand for minimally invasive surgeries has increased patient traffic in Cath Labs, placing considerable strain on resources, personnel, and procedural efficiency. Challenges such as scheduling conflicts, resource utilization, and staff fatigue [4] have made effective Cath Lab management more critical than ever. Automated surgical workflow analysis presents significant potential to address these issues by delivering real-time insights into procedural phases, enabling enhanced resource management and informed scheduling.

Traditional workflow analysis methods, such as manual annotation, are labour-intensive and impractical for real-time applications. Recent advancements in machine learning (ML) and computer vision (CV) offer scalable alternatives, automating workflow analysis through visual data. However, current approaches face challenges in critical clinical environments due to their lack of

[1]Christos Spiliadis is with the Department of Microelectronics, Delft University of Technology, 2628 CD Delft, The Netherlands c.spiliadis@student.tudelft.nl
[2]Yiheng Chang is with the Department of Microelectronics, Delft University of Technology, 2628 CD Delft, The Netherlands
[3]Justin Dauwels is with the Department of Microelectronics, Delft University of Technology, 2628 CD Delft, The Netherlands j.h.g.dauwels@tudelft.nl
[4]Chavdar Bachvarov is with the Department of IGT-S Innovation, Philips Medical Systems, 5684 PC Best, The Netherlands chavdar.bachvarov@philips.com
[5]John J. van den Dobbelsteen is with the Department of Mechanical Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands j.j.vandendobbelsteen@tudelft.nl
[6]Benno H. W. Hendriks is with the Department of IGT-S Innovation, Philips Medical Systems, 5684 PC Best, The Netherlands b.h.w.hendriks@tudelft.nl
[7]Maarten Van der Elst is with Reinier de Graaf Groep, Delft, The Netherlands m.vanderelst@tudelft.nl
[8] Dr. Markku Eskola is with HH Tampere, Tampere, Finland markku.eskola@sydansairaala.fi

interpretability and generalizability. Most studies rely on endoscopic cameras or black-box deep learning models, limiting their transparency and adaptability. Furthermore, no existing solutions using external cameras have demonstrated the ability to generalize across datasets from different hospitals, leaving a critical gap in the literature.

To address these limitations, this research focuses on ceiling-mounted cameras for input. It introduces a two-stage model combining YOLOv8 object detection and a Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) framework. This approach emphasizes explainability and generalizability, enabling precise detection of surgical workflow phases and their transitions. The proposed model has been validated using datasets from two distinct Cath Lab environments, demonstrating its ability to adapt to different hospital workflows.

In addition to addressing workflow efficiency, the study highlights the need for better Cath Lab management to mitigate personnel risks. Frequent exposure to ionizing radiation, the physical demands of long hours in protective gear, and high-pressure environments contribute to long-term health challenges for staff [4]. Workflow analysis can improve team coordination, reduce inefficiencies, and enhance safety for medical teams and patients.

The primary contributions of this study are as follows:

1) A novel, explainable two-stage model integrating YOLOv8 and GMM-HMM for phase inference.
2) Validation of the model's generalizability across two hospital datasets.
3) Demonstration of the practical application of real-time automated surgical workflow analysis to improve Cath Lab operations.

This paper will discuss the proposed approach's methodology, results, and implications, offering insights into its potential to enhance procedural efficiency and safety in Cath Labs.

## II. METHODS

This study proposes a two-stage model for surgical workflow analysis in Cath Labs. The model integrates object detection and temporal phase inference to detect and classify workflow phases accurately. It utilizes ceiling-mounted cameras to capture video data, which is processed in two stages to produce interpretable and generalizable insights.

### A. First Stage

In the first stage, the YOLOv8 [5] model was applied to three different camera angles to extract spatial features essential for the second stage of temporal phase inference. These camera angles were selected per dataset to optimize visibility and minimize occlusions caused by medical equipment in the Cath Lab. The primary camera angle generally provided a direct view of the operating table and cath lab door, ensuring continuous patient movement tracking. The secondary camera offered an additional perspective that helped distinguish object interactions, while the tertiary camera complemented the spatial coverage, particularly capturing patient entry and exit events.

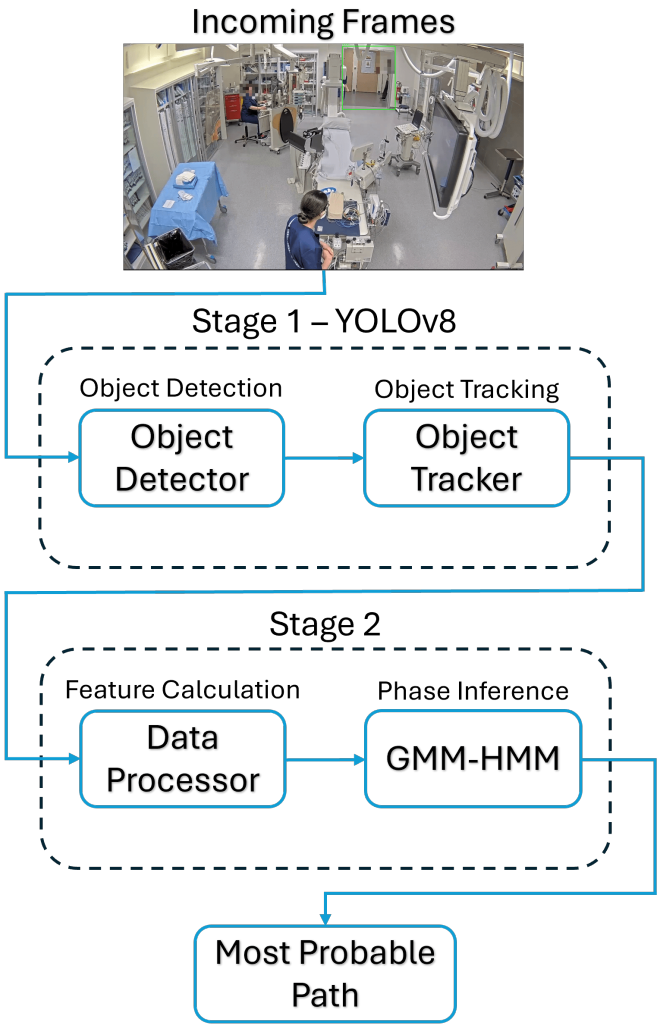


Fig. 3. Entire model layout

Initially, the YOLOv8 model was trained on the Renier de Graaf Groep (RdGG) object dataset. This dataset was curated to ensure high-quality annotations of key objects within the Cath Lab environment, including patients and doors, which were crucial for workflow phase classification. The training process followed the methodology outlined in [3], where it previously optimized YOLOv8 for detecting objects in the Cath Lab RdGG Hospital. The best-performing weights from this initial training phase were used when the complete model was applied to RdGG procedures.

To extend the generalizability of the object detector across different hospital environments, the model was then re-trained on the object detection dataset from Heart Hospital Tampere (HH Tampere). This re-training phase allowed the model to adapt to variations in light-

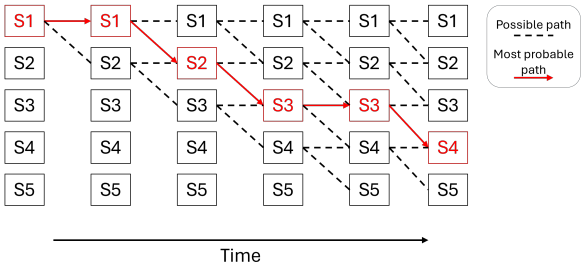


Fig. 4. Illustration of the Viterbi decoding process for a left to right HMM

ing, equipment positioning, patient attire, and procedural workflows unique to the HH Tampere setting. The dataset consisted of multiple procedure videos recorded with ceiling-mounted cameras, capturing diverse patient movements and procedural contexts in 69237 frames. The training methodology incorporated transfer learning techniques, where the previously fine-tuned weights from the RdGG training phases were used as an initialization point, thereby reducing training time and improving model stability [6].

## B. Second Stage

The second stage of the model employs a Gaussian Mixture Model - Hidden Markov Model (GMM-HMM) framework to infer the temporal phase sequence of a surgical procedure. This approach integrates spatial features extracted by YOLOv8 with temporal dependencies to classify each second of the procedure into one of five predefined workflow phases: Waiting, Patient Entering, Patient Laying, Patient Exiting, and End, also visible in Figure 5. The combination of GMM and HMM allows the model to accommodate intra-phase variability while enforcing structured phase transitions.

The Gaussian Mixture Model (GMM) component is responsible for modelling the probability distributions of spatial features within each workflow phase. Since features exhibit multimodal distributions due to procedural variations, the GMM provides a flexible probabilistic representation by approximating these variations with multiple Gaussian components [7]. Each phase-feature combination is assigned a distinct GMM, with its parameters estimated using the Expectation-Maximization (EM) algorithm. The EM algorithm iteratively refines the mean vectors, covariance matrices, and mixture weights, allowing the model to capture subtle differences in spatial characteristics across different phases [8], [9]. Only correlated features are concatenated to ensure meaningful representation, preventing redundant or misleading information from affecting the model's ability to differentiate between workflow stages [10].

The Hidden Markov Model (HMM) governs the sequential nature of phase transitions using a transition matrix that defines the probability of moving from one phase to another [11]. This transition matrix is derived from annotated training data and encodes logical constraints to prevent unrealistic transitions. For example, a direct transition from Waiting to Laying is disallowed, as a patient must first enter the Cath Lab. Conversely, transitions such as Entering to Laying are highly probable, representing common procedural progressions. The left-to-right topology of the transition matrix ensures that phases progress in a natural order, mimicking real-world surgical workflows and preventing illogical regressions between phases.

The model employs the Viterbi algorithm to determine the most probable sequence of workflow phases given an observation sequence. This dynamic programming method iteratively calculates the highest probability of reaching each phase at every time step, leveraging previously computed probabilities to determine the optimal phase sequence efficiently [11]. The Viterbi algorithm also incorporates backtracking to reconstruct the most likely phase sequence, ensuring the inferred workflow is accurate and interpretable.

The Viterbi algorithm operates through three main steps: initialization, recursion, and termination. During initialization, the algorithm assigns initial probabilities to each state. The recursion step iteratively computes the highest probability of reaching each state at each time step, considering transition probabilities and observation likelihoods, which in this case come from both the discrete features as well as the membership probabilities calculated from the Gaussian mixtures. Finally, in the termination step, the algorithm selects the most probable final state and backtracks through stored values to reconstruct the optimal phase sequence.

1) Initialization:

$$v_1(j) = \pi_j b_j(o_1), \quad 1 \le j \le N \tag{1}$$

$$bt_1(j) = 0, \quad 1 \le j \le N \tag{2}$$

2) Recursion:

$$v_t(j) = \max_{1 \le i \le N} v_{t-1}(i) a_{ij} b_j(o_t), \quad 1 \le j \le N, \quad 1 < t \le T \tag{3}$$

$$bt_t(j) = \arg\max_{1 \le i \le N} v_{t-1}(i) a_{ij} b_j(o_t), \quad 1 \le j \le N, \quad 1 < t \le T \tag{4}$$

3) Termination:

$$P^* = \max_{1 \le i \le N} v_T(i) \tag{5}$$

$$q_T^* = \arg\max_{1 \le i \le N} v_T(i) \tag{6}$$

These equations illustrate the core steps of the Viterbi algorithm. The recursion step ensures that the most probable sequence is selected by maximizing over previous state probabilities ($v_{t-1}$), transition probabilities ($\alpha_{ij}$), and observation likelihoods ($b_j(o_t)$). The backtracking step ($bt$) reconstructs the optimal sequence by

tracing back through stored indices, ensuring accurate workflow phase detection.

All probability calculations are performed on a logarithmic scale to prevent numerical instability, mitigating the risk of underflow errors that could arise from multiplying small probability values over long sequences.

The model incorporates an online inference mechanism based on a modified Viterbi algorithm for real-time phase inference. Unlike traditional offline inference, where phase detection is performed after an entire procedure has been processed, this approach allows the model to make incremental phase predictions as new data arrives. A dynamic buffering mechanism ensures stability in real-time predictions, where a sliding window maintains a history of phase predictions. This mechanism prevents abrupt phase transitions by validating them across multiple frames before finalizing a decision. The buffer size is dynamically adjusted based on the stability of past predictions, optimizing the balance between responsiveness and accuracy. This online inference method allows the model to be deployed in real-time surgical settings, providing immediate insights into workflow phases while maintaining robustness against transient detection errors or occlusions.

Creating features from the bounding boxes remains consistent across datasets and is performed by the data processor shown in Figure 3, ensuring model generalizability to different hospital environments. Regardless of dataset variations, spatial and temporal features are derived following the same methodology, allowing seamless model adaptation to new clinical settings. The extracted features are categorized into continuous and discrete types, each critical in workflow phase detection.

Continuous features model gradual transitions and dynamic aspects of the procedure. These include the patient's bounding box coordinates, the patient-door distance, the Intersection over Union (IoU) between detected objects, and the centroid position of the patient's bounding box from multiple camera views. These features provide nuanced information about spatial relationships and movements, ensuring smooth tracking of the procedure's progression.

Discrete features, however, capture abrupt changes in the workflow that indicate phase transitions. Examples include patient visibility (whether the patient is in the field of view), final patient exit confirmation (differentiating between temporary absences and the end of the procedure), and patient existence confirmation (validating transitions between major workflow phases). These features act as structural constraints that reinforce the logical consistency of the phase inference process.

The model achieves a balanced approach to phase detection by combining continuous and discrete features. Continuous features provide fine-grained motion analysis, while discrete features ensure logical sequencing and stability. This dual-feature integration allows the model to accurately capture both gradual procedural changes and sudden transitions, leading to robust and interpretable surgical workflow analysis.

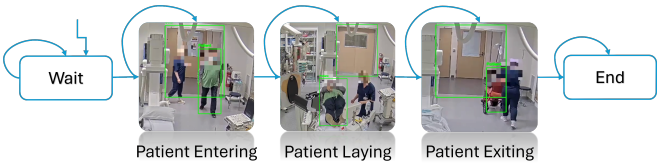


Fig. 5. The five phases that are of importance to this model.

## III. DATASETS

The datasets used in this study were acquired from two different hospital environments, ensuring variability in workflow dynamics and procedural execution. These datasets were recorded in Catheterization Laboratories (Cath Labs) equipped with ceiling-mounted cameras, capturing surgical workflow data from different perspectives. All procedures were recorded with the approval of the Medical Ethical Testing Committee (METC) and the informed consent of both staff and patients. Both datasets utilized in this research are proprietary and provided by Philips.

The first dataset, consisting of 55 procedures, was collected at Renier de Graaf Hospital (RdGG) in the Netherlands. It primarily consists of video recordings of diagnostic coronary angiography procedures. The recorded procedures follow a structured workflow with minimal variations, as patient attire and procedural execution remain relatively consistent. The dataset is characterized by clear visibility of patient entry, positioning on the operating table, and subsequent procedural steps. This structured nature makes it well-suited for training object detection and phase inference models with controlled variations.

The second dataset, which features diagnostic and therapeutic interventions, was acquired from HH Tampere in Finland and consists of 15 procedures. Unlike the RdGG dataset, this dataset exhibits more significant procedural variability due to different patient transport methods, varied patient attire, and a broader range of interventional techniques. This dataset includes percutaneous coronary interventions (PCI) alongside diagnostic procedures, making it more diverse in terms of workflow complexity. The increased variability in environmental factors and workflow execution presents a more challenging testbed for assessing the generalizability of the proposed model.

The datasets were processed to ensure each procedure was represented as one video per angle, with all angles aligned to maintain correct timing. Phase annotations were performed by non-medical personnel, as medical expertise was not required to recognize these straightforward phases. The annotation process involved finding the four transition times, and once all four transitions were identified, the calculated features could be split

into five phases for training the GMM-HMM. The combination of these datasets allows for the validation of the model across distinct hospital settings, enhancing its adaptability and robustness in real-world applications.

## IV. EXPERIMENTS AND RESULTS

The experimental results illustrate the impact of varying Gaussian components and transition window lengths on model performance across the RdGG and HH Tampere datasets. Figure 6 and 7 show that for the RdGG dataset, performance peaks around 20 components, except for 23 components, which had a slightly higher average F1 score but inconsistent results across procedures and were discarded. The HH Tampere dataset follows a similar trend, where F1 scores improve up to 15 components, stabilize around 20, and show minor fluctuations beyond that. Additionally, Figure 7 demonstrates that computation time per fold increases almost linearly with the number of components, with the RdGG dataset requiring more processing time due to its larger number of procedures per fold. Based on these observations, 20 components were selected for each GMM, balancing performance with computational efficiency and ensuring real-time applicability.

Another crucial factor in model performance was the choice of transition window length [12]. Figure 8 illustrates how F1 scores increase with window size, with diminishing returns beyond a certain threshold. The RdGG dataset consistently achieved higher F1 scores than the HH Tampere dataset, stabilizing at approximately $\pm 6$ seconds. Meanwhile, the HH Tampere dataset saw the most substantial improvement between $\pm 1$ and $\pm 3$ seconds due to its shorter wait and end phases, which otherwise led to stricter penalties for timing mismatches. As a result, a $\pm 3$ second transition window was selected to optimize precision and recall, minimizing minor timing discrepancies while ensuring meaningful transition alignment.

The RdGG dataset exhibited consistently high precision, recall, and F1 scores across all phases, particularly in longer, well-defined phases like Lay and End. The structured nature of RdGG procedures contributed to this performance, as seen in Table I. However, challenges emerged in transitional phases, such as Enter and Exit, where brief durations led to occasional timing misalignments. Some patients exhibited non-standard behaviours, such as pausing before fully entering or exiting the Cath Lab, sometimes resulting in false transitions. Despite these minor misclassifications, Figure 10 demonstrates strong precision scores across all phases, reinforcing the model's reliability in structured surgical workflows. Overall, the model achieved an average F1 score of 95.4%, highlighting the model's ability to perform phase inference in surgical environments accurately.

In contrast, the HH Tampere dataset presented more significant procedural variability, making phase detection more challenging. Enter and Exit phases had increased misclassification rates due to patient movement and positioning differences, as shown in Figure 9. Some procedures involved patients sitting for a prolonged time before lying down, a behaviour less common in RdGG. This variation led to difficulties distinguishing Enter from Lay phases, as reflected in lower recall scores for the Enter phase in Figure 11. Despite these challenges, the model achieved an overall F1 score of 95.2%, demonstrating its ability to generalize across hospital environments. The increased flexibility required for HH Tampere procedures highlights the importance of adaptable phase inference methods.

A cross-dataset analysis reveals key differences influencing phase detection performance. The RdGG dataset, characterized by a structured workflow with distinct, longer phases, consistently produced higher accuracy across all phases. Conversely, the HH Tampere dataset exhibited greater variability in patient movement and procedural execution, making transitional phases more challenging to detect. Dataset size also played a role in performance variation, with RdGG's larger dataset allowing for better parameter estimation and more stable phase transitions. The impact of dataset complexity is evident in the transition window analysis, whereas the improvements were minor on the RdGG dataset. In contrast, the HH Tampere dataset showed more pronounced improvements between $\pm 1$ and $\pm 3$ seconds.

A real-time demonstrator was implemented and tested using two procedures from the RdGG dataset to evaluate the real-time applicability of the model. This evaluation aimed to compare the performance of the real-time Viterbi algorithm with the classical Viterbi approach, which processes the entire observation sequence before determining the most probable phase path. The real-time implementation processed incoming video frames sequentially, using an adjustable buffer to stabilize phase transitions. The comparative analysis showed that the real-time Viterbi algorithm produced identical phase sequences to the classical version, confirming that the model maintains accuracy even in live surgical environments. However, it was only tested on two procedures since some procedures showed issues with the model stability, particularly in cases with long inactivity periods when the patient is not visible to any camera. For instance, in the 'Patient Laying' phase, the model would prematurely move to the 'Patient Exiting' and 'End' phases as the patient was not visible, and since a camera would not see the patient for a long while, the buffer would not get enough inputs in time to recognize this path as an unstable path.

## V. DISCUSSION & CONCLUSIONS

The results of this study highlight the potential of an explainable and generalizable model for surgical workflow analysis in Cath Labs. Unlike black-box deep learning models that provide limited transparency, the proposed two-stage approach effectively separates spatial feature

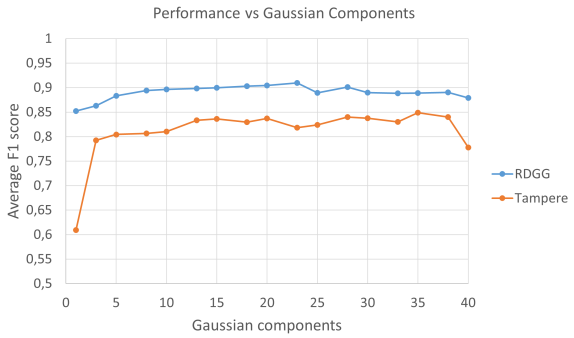


Fig. 6. Average F1 score versus the number of Gaussian components present in each GMM for the RdGG and HH Tampere datasets.

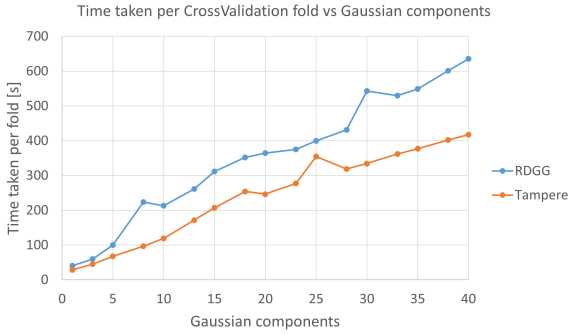


Fig. 7. Average time taken per fold versus the number of Gaussian components present in each GMM for the RdGG and HH Tampere datasets.

| | Precision | | Recall | | F1 | |
|---|---|---|---|---|---|---|
| | RdGG | Tamp. | RdGG | Tamp. | RdGG | Tamp. |
| Wait | 0.997 | 1 | 0.995 | 0.966 | 0.996 | 0.977 |
| Enter | 0.934 | 0.806 | 0.979 | 0.943 | 0.943 | 0.831 |
| Lay | 1 | 0.995 | 0.995 | 0.998 | 0.997 | 0.996 |
| Exit | 0.842 | 0.997 | 0.956 | 0.955 | 0.856 | 0.965 |
| End | 0.986 | 1 | 0.981 | 1 | 0.979 | 1 |
| Avg. | 0.952 | 0.957 | 0.981 | 0.972 | 0.954 | 0.952 |

TABLE I
F1, Precision, and Recall exact scores for the RdGG and HH Tampere datasets

extraction from temporal phase inference, ensuring a more interpretable decision-making process. The combination of YOLOv8 for object detection and GMM-HMM for temporal modelling provides a structured, probabilistic framework capable of capturing workflow dynamics while maintaining real-time feasibility.

One of the model's key strengths is its robust generalization across hospitals. The model was validated in two distinct clinical environments by leveraging datasets from RdGG in the Netherlands and HH Tampere in Finland. Despite differences in procedural workflows, patient transport methods, and attire, the system maintained high accuracy, achieving F1-scores of 95.4% for RdGG and 95.2% for HH Tampere (Table I). This demonstrates

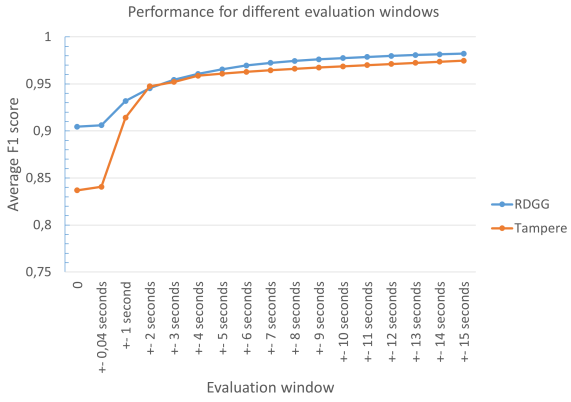


Fig. 8. Average F1 score versus transition window size for the RdGG and HH Tampere datasets

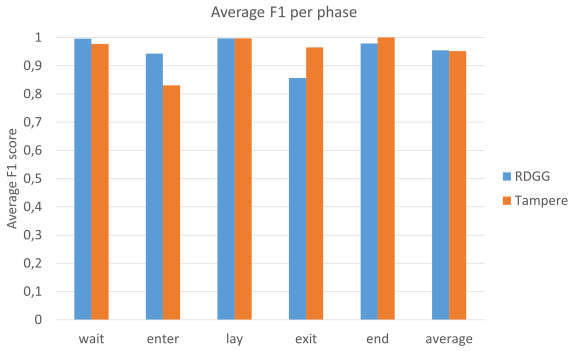


Fig. 9. F1 scores per phase

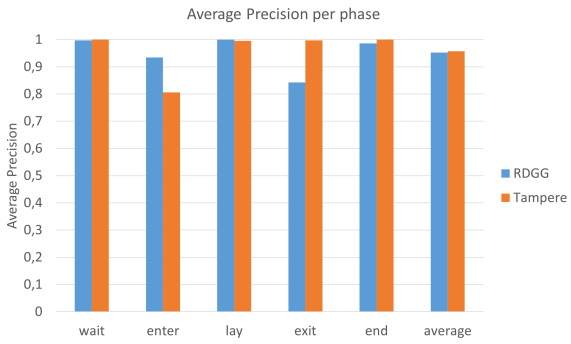


Fig. 10. Precision scores per phase

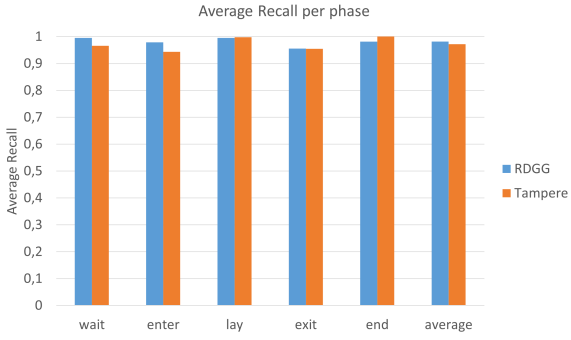


Fig. 11. Recall scores per phase

that the method successfully adapts to varied hospital settings, making it a scalable solution for broader clinical adoption.

Moreover, the structured phase-specific feature extraction method enhances interpretability. Continuous features, such as patient bounding box coordinates and the distance between the patient and the door, help model gradual transitions, while discrete features, such as final patient exit confirmation, ensure phase alignment. The phase-specific analysis (Figures 9-11) reveals that longer and more structured phases, such as Lay and End, are detected with high accuracy, while shorter transitional phases, such as Enter and Exit, present more challenges due to procedural variability.

Despite its strengths, the model has certain limitations. First, edge cases, such as patient occlusions by medical staff or non-standard entry/exit behaviours, occasionally lead to misclassifications. The HH Tampere dataset, in particular, exhibited more procedural variability, making phase boundaries more ambiguous. Additionally, while the transition matrix and Viterbi decoding enforce logical workflow constraints, the model still depends on the quality of its training data. Expanding the dataset with more procedures and hospital environments would improve robustness and minimize edge case errors.

The real-time demonstrator confirmed the practical feasibility of the proposed model, showing that stable phase predictions can be integrated into workflow-monitoring tools. However, challenges remain, particularly prolonged patient inactivity and occlusions, which can disrupt phase stability. Refining the buffering mechanism and adding contextual features should improve robustness, while broader real-time testing will further validate reliability. For occlusions, adjusting current ceiling cameras to give complementary views reduces blind spots, and fusing synchronized multi-view footage in a lightweight 3-D reconstruction can recover obstructed regions, maintaining consistent tracking during critical moments.

From a practical perspective, the model's accurate phase detection can streamline resource allocation and minimize idle times in the Cath Lab. By recognizing the current phase, staff can anticipate upcoming tasks more precisely, reducing scheduling gaps and potentially increasing patient throughput. In addition, metrics on phase durations and resource usage facilitate feedback for training programs and workflow optimization.

Future work will focus on enhancing the model's adaptability in clinical settings. One promising direction is expanding object detection capabilities beyond patients and doors to include medical instruments and staff, enabling finer-grained workflow analysis. Additionally, integrating multimodal data sources, such as X-ray usage patterns, angiography signals, or audio cues, could improve phase inference granularity and increase the model's decision-making accuracy. Another avenue of research is adapting the model to other surgical environments, such as operating rooms for general or orthopaedic surgery, extending its utility beyond the Cath Lab.

In conclusion, this study presents a scalable and explainable approach to surgical workflow analysis, addressing the dual challenges of interpretability and generalizability. Integrating YOLOv8 with GMM-HMM ensures a structured and robust phase inference framework while maintaining real-time feasibility. The results demonstrate that phase transitions can be accurately detected in different hospital settings, contributing to improved operational efficiency, resource allocation, and procedural safety. As advancements in AI-driven healthcare continue, this model serves as a foundation for future intelligent surgical workflow monitoring systems.

## References


[1] "Cardiovascular diseases". World Health Organisation. Accessed: April 9, 2025. [Online.] Available: https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1.

[2] L. Ilcheva et al., "Beyond Conventional Operations: Embracing the Era of Contemporary Minimally Invasive Cardiac Surgery". Journal of Clinical Medicine, vol. 12, no. 23, p. 7210, 2023, doi: https://doi.org/10.3390/jcm12237210.

[3] Y. Chang, "Video-based Event Detection in Catheterization Lab", M.Sc. thesis, Microelectronics Dept., TU Delft, Delft, 2023.

[4] M. G. Andreassi, E. Piccaluga, G. Guagliumi, M. Del Greco, F. Gaita, and E. Picano, "Occupational Health Risks in Cardiac Catheterization Laboratory Workers", Circulation: Cardiovascular Interventions, vol. 9, no. 4, Apr. 2016, doi: https://doi.org/10.1161/circinterventions.115.003273.

[5] "Ultralytics YOLOv8 | State-of-the-Art Vision AI," www.ultralytics.com. Available: https://www.ultralytics.com/yolo

[6] D. Soekhoe, van, and A. Plaat, "On the impact of data set size in transfer learning using deep neural networks," H. Boström, A. Knobbe, C. Soares, and P. Papapetrou, Advances in Intelligent Data Analysis XV. IDA 2016. Lecture Notes in Computer Science, vol 9897, Springer International Publishing, 2016, pp. 50–60.

[7] S. Theodoridis, Machine learning: a Bayesian and optimization perspective. London, San Diego: Elsevier: Academic Press, 2020.

[8] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data Via the EM Algorithm," Journal of the Royal Statistical Society: Series B (Methodological), vol. 39, no. 1, pp. 1–22, Sep. 1977, doi: https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

[9] G. J. Mclachlan, D. Peel, and I. Netlibrary, Finite Mixture Models. New York: Wiley, 2000.

[10] J. Lei, J. Zhang, G. Li, Q. Guo, and D. Tu, "Continuous action segmentation and recognition using hybrid convolutional neural network-hidden Markov model," IET Computer Vision, vol. 10, no. 6, pp. 537–544, Sep. 2016, doi: https://doi.org/10.1049/iet-cvi.2015.0408

[11] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proceedings of the IEEE, vol. 77, no. 2, pp. 257–286, 1989, doi: https://doi.org/10.1109/5.18626

[12] O. Dergachyova, D. Bouget, A. Huaulmé, X. Morandi, and P. Jannin, "Automatic data-driven real-time segmentation and recognition of surgical workflow," International Journal of Computer Assisted Radiology and Surgery, vol. 11, no. 6, pp. 1081–1089, Mar. 2016, doi: https://doi.org/10.1007/s11548-016-1371-x