REDUCING WORKFLOW MODEL COMPLEXITY FOR AN AUTOMATIC WORKFLOW DISCOVERY ALGORITHM

By

JUHI TRIPATHI

A thesis submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Electrical and Computer Engineering

Written under the direction of

Dr. Ivan Marsic

And approved by

____

____

____

____

New Brunswick, New Jersey

October 2019

ABSTRACT OF THE THESIS

REDUCING WORKFLOW MODEL COMPLEXITY FOR AN AUTOMATIC WORKFLOW DISCOVERY ALGORITHM

by JUHI TRIPATHI

Thesis Director:

Dr. Ivan Marsic

With the increasing number of adverse patient outcomes and deaths due to medical errors, process mining techniques have been used to build automatic workflow discovery algorithms to assist teams during the treatment procedure, better understand the treatment practice and potentially help improve patient outcomes. These automatic workflow discovery algorithms use event logs that has all the information of the events that are executed in the treatment procedure, to extract a process model. Automatic workflow discovery algorithms are now used to analyze the trauma resuscitation process and improvise the patient outcomes. The widespread use of process mining techniques to discover and analyze workflows in healthcare, has motivated this study that has analyzed, how varying a parameter that contributes to the workflow discovery algorithm changes the workflow model, discusses the reasons behind those changes in the discovered workflow, presents the optimum value of the parameter that produces best results and backs up the theoretical discussion with numerical results. We use an expert-based model that is derived from hand drawn model by medical experts after multiple revisions, that serves as the ground truth and to compare the accuracy of our workflow model. Based on the comparison of the workflow model with the expert model, the variations observed could be classified into the categories : an incorrect addition of

treatment activity, a redundant addition of treatment activity or a correct addition of treatment activity to the workflow. This work is an attempt in the direction to make the workflow model generated using the workflow algorithm, more accurate and exhibits least complexity, thereby making it easy to comprehend. The workflow discovery algorithm used for this work has two phases : (1) Construction of an event sequence of consensus activities that has the occurrence probability of more than the predefined threshold. This occurrence probability is determined using the event logs. This phase ensures that the workflow model comprises of the frequent activities that have appeared in the event logs. (2) Inclusion of non-consensus activities, that are common but dispersed between consensus activities in the workflow. This inclusion is done after multiple iterations through consensus activities to determine the activities that are interleaved between consensus activities and when considering a window of consecutive consensus activities, the combined probability of occurrence of the non-consensus is more than the predefined threshold.

This study explores one of the contributing hyperparameters of the algorithm, to infer its effect on the workflow model. The analysis falls in agreement with the underlying notion behind this study, that varying the hyperparameter of workflow discovery algorithms results in varying complexity of the workflow model.

Acknowledgement

I would like to express sincere gratitude to my thesis advisor, Dr. Ivan Marsic, faculty in Electrical and Computer department, at Rutgers, The State University of New Jersey, for imparting his knowledge and expertise, that has always helped me work in the right direction. Dr. Marsic has always been motivating and has guided me at every step to make sure that the consistent quality of the work is maintained throughout my work. Without his guidance and help, this work would have not been possible.

I am grateful to Dr. Zoran Gajic, graduate director of the ECE department at Rutgers, The State University of New Jersey for his constant support and guidance throughout my time at Rutgers. Thank you for giving me the opportunity to be a part of this experience at Rutgers, that has helped me grow and learn more than I ever had in such a short span of time. I thank the administrative staff for answering all my queries and making sure that I stay compliant.

I would also like to thank the distinguished members of committee, Dr. Maria Striki and Dr. Yingying Chen, for taking out time from their busy schedules to be a part of my defense committee.

Lastly, I would like to extend my thanks and appreciation to my family and friends who have stood by me through every step of my journey and believed in me.

Table of contents

List of tables

1. Introduction

The number of deaths in the United States, caused by errors in medical practice has seen a rapid increase over the past decade [1]. According to the Institute of Medicine, Medical error is defined as “failure to complete a planned action as intended or the use of a wrong plan to achieve an aim” and is now the third leading cause of death in the United States [2]. While half of the preventable deaths are pertinent to the errors that are instigated during the preliminary resuscitation of the treatment procedure [3][4], there has been multiple studies and attempts in the direction to plummet the adverse outcomes caused by medical errors - Computer Aided decision support being one of them, that assures a more precision medicine while being cost effective [5]. In the direction to develop such a system, one of the most crucial steps is the recognition of deviations in the treatment activities involved in a medical procedure, to avoid any sort of adverse outcomes. For a trauma resuscitation, based on the identified individual parameters (such as injuries), a protocol compliant action plan is developed and executed. The conventional workflow models were developed after multiple revisions by experts, till an agreement is reached to confirm the safety and reliability of the workflow for the initial assessment and management of the trauma patient. Computer aided decision support system generates such a workflow using data mining techniques, over a given dataset of activity traces for a medical process.

2. Problem Statement

The interpretability of a medical workflow model is one of the most important factors in the discovery and analysis of a workflow. In order to avoid the formation of a complex model, that is difficult to comprehend, the proposed workflow discovery algorithm [6] accomplished the goal of presenting a better interpretable workflow model as compared to the baseline model, discussed in the paper. The workflow discovery algorithm [6] has the advantages of producing sequential, and no loop (only left-to-right structure) models, that refined the interpretability of the produced workflow to a large extent. While the preliminary results testified the proposed algorithm, a comprehensive analysis and evaluation, could further lead to a more refined model.

The goal of this work is to study and optimize the discovered workflow model that is derived for a given dataset of activity traces for a medical process (trauma resuscitation), using data mining techniques. The key challenge is the “noise” present in activity traces, which is caused by process flexibility (so activities do not need to be performed in a strict order), providers’ personal preferences, or human errors. This “noise” clutters the process model and makes it needlessly complex and difficult to distinguish the essential from accidental elements of the discovered workflow.

3. Workflow Discovery Algorithm and Parameters Used

The 2 major steps in the discovery of workflow model involve :

(i) Formation of an alignment matrix

(ii) Deriving a workflow model from the alignment.

While deriving the sequence from the alignment matrix, an important hyperparameter that plays a crucial role is the ‘’Span’’. Before discussing span, it is important to understand what is ‘Consensus Activity’. The trace alignment matrix has columns that corresponds to activities that have occurrence probability more than a predefined threshold (□) - these columns are referred to as the Consensus columns, and the remaining columns are known as Non-consensus columns. Consensus columns together form the backbone structure of the workflow model and gives us the information about all the necessary steps in a medical workflow. Across a given span of columns, if a non-consensus activity occurs with frequency more than the predefined threshold - it is referred to as common but dispersed activity. Thus, span is the maximum number of consecutive consensus columns across which common but dispersed activities are spread.

Varying the value of span, changes the number of consecutive consensus columns across which, we find the common but dispersed activities. If the value of span is lesser than the ideal value (the optimum value of span - discussed in the following sections), the number of consecutive consensus columns will be lesser and the frequency of occurrence of crucial activities across the narrow window of consensus column may not be more than the threshold, which could result in a workflow model that is deprived of these crucial activities. Similarly, if the value of span is more than the optimum value of span, the number of consecutive consensus columns will be more and the frequency of occurrence of non-crucial activities across the broad window of consensus column may be more than the threshold, which could result in a workflow model that comprises of these non-crucial activities. Therefore, as we change the value of span, the common but dispersed activities in the workflow model vary. Along these lines of varying the span, we explore the idea of optimum value of span, at which the workflow model generated using workflow discovery algorithm [6], is most accurate. An important factor to consider the accuracy of workflow model is the expert model, that is the workflow model derived from the hand-drawn models by experts. For now, we consider the hand drawn workflow model, provided by experts (referred to as expert model in this study) as the “ground truth”, but we will always watch for potential problems where a disagreement between the discovered workflow model and expert model needs to be discussed with experts and could result in iterative revisions of the expert model, till we reach an agreement.

4. Hypothesis

The use of span can greatly reduce the model complexity in complex processes. This hypothesis was first proposed in the paper "Discovering Interpretable Medical Workflow Models" [6]. The aim of this study is to analyze span values and justify that, varying span could actually result in increase or decrease of a model interpretability, complexity and accuracy.

Experimental results and analysis carried out in the direction of the model optimization are further elaborated to highlight that an optimum value of Span for a given workflow model produces the optimum result. An optimum model here refers to the model that has minimal 'noise' and exhibits closest resemblance to the hand drawn workflow models, provided by experts.

5. Description of Dataset Used

For a medical procedure of Intubation, we will be considering two samples of intubation dataset, one is a smaller dataset that has 168 treatment activities and the other is a larger dataset of 1240 treatment activities. The treatment activities are mentioned corresponding to their case IDs, where each case ID refers to individual patient case. For a single case ID, all the treatment activities carried out, form an activity trace. In other words, we have multiple activity traces, where each activity trace has a variable number of treatment activities. Since, both the datasets used are for the same procedure of Intubation, and only differ in the number of treatment activities, the dataset that is henceforth referred to as the smaller intubation dataset is the one with lesser number of treatment activities than the other, and the one that is henceforth referred to as the larger intubation dataset is the one with more number of treatment activities than the smaller dataset. The different parameters of discussed datasets are given below -

- The number of different activity types in both datasets is 15.
- The number of activity traces in smaller intubation dataset is 7, and in larger intubation dataset is 101.
- Average length of activity trace for smaller intubation dataset is 13, and for larger intubation dataset is 12.

In the following sections, we proceed with the detailed analysis for the smaller intubation dataset, followed by the detailed analysis of larger intubation dataset.

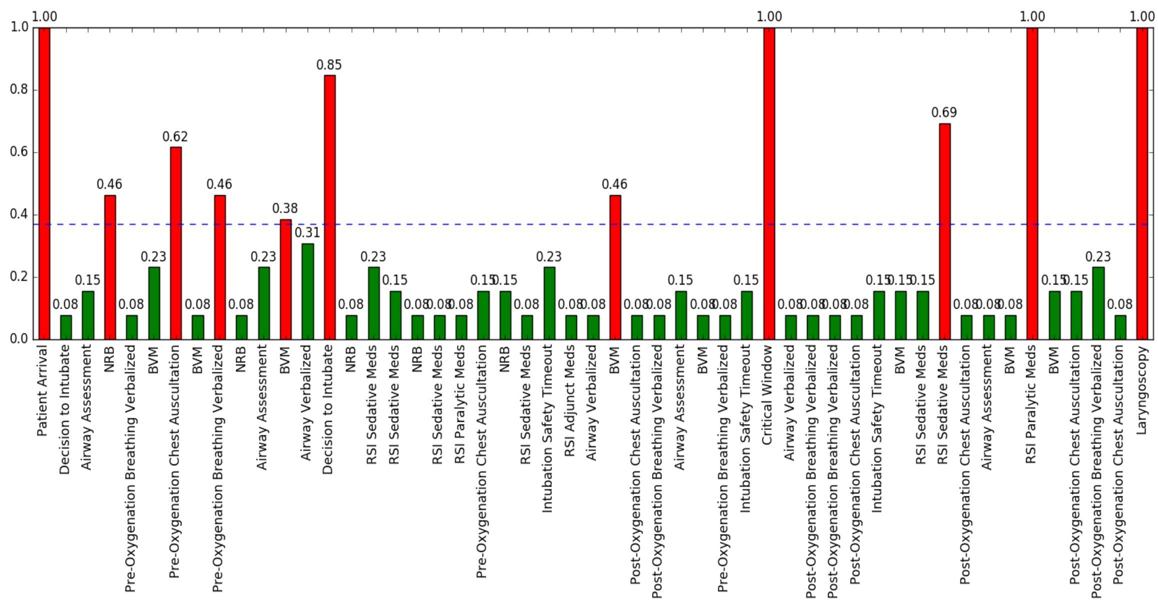


Figure 1 : Treatment activities vs probability of occurrences plot for smaller intubation dataset

Figure 1 shows the treatment activities for intubation, plotted against their probability of occurrence for the smaller intubation dataset. These activities are ordered by their appearance in the alignment matrix. To illustrate, the activity traces may have different orders of the treatment activities. Every activity trace starts with the activity “Patient Arrival”, and ends with the activity “Laryngoscopy”, therefore these activities don't appear anywhere else in the plot except for the first and last columns. However, other activities might be performed in different orders for different activity traces, and thus, appear in the order of their appearance in alignment matrix. The threshold (□) is set to 0.37, thus giving all the activities with probability of occurrence greater than 0.37 as consensus activities - red colored, and the rest of them as non-consensus activities - green colored. It is clear from the diagram that the activities in red, form the backbone of workflow model, but amongst the non-consensus columns, consider the activity ‘Airway Assessment’ that is interleaved between consensus activities, such that each time its probability of occurrence is less than threshold, but across a span of consecutive consensus columns it occurs multiple times and the probability of occurrence adds up to more than the threshold. To be precise, for a span of 4 consecutive consensus activities, ‘Airway Assessment’ occurs twice with probability of occurrence 0.15 and 0.23 respectively, thus adding up to a value of 0.38 and crossing the threshold. Therefore, for the span of 4 consecutive consensus activities, ‘Airway Assessment’ is added to the workflow. The experimental results shared in the following sections (span analysis for 4 consecutive consensus activities) fall in agreement with the above analysis of having ‘Airway Assessment’ introduced in the workflow at span of 4 consecutive consensus activities, and not for any lesser values of span. Thus, Airway Assessment could be seen as an important treatment activity that wasn't included in the workflow models generated for span values 0-3 consecutive consensus activities, but for span value of 4 consecutive consensus activities, the activity is added to the workflow, bringing it closer to the actual workflow.

6. Analysis for Smaller Intubation Dataset

6.1. Analysis Using Smaller Intubation Dataset for Span of 0 consecutive consensus activity

For Span of 0 consecutive consensus activity, it corresponds to the case where we go through 0 consecutive consensus columns to determine the common but dispersed activities. This necessarily means that we would have no such common but dispersed activity, and the workflow model will only comprise of the backbone formed by the consensus activities.

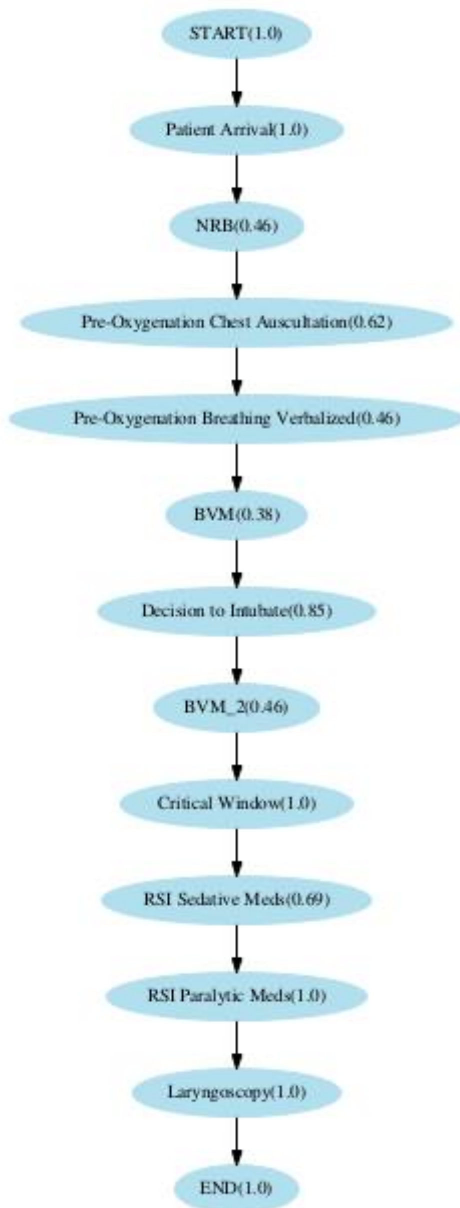


Figure 2 : Workflow model for smaller dataset at span of 0 consecutive consensus activity

The workflow model generated for the span of 0 consecutive consensus activity, is shown in figure 2. Conforming to our analysis in the previous paragraph, the model contains only the backbone structure with no common but dispersed activity. Figure 2 shows the linear workflow model, with no branches because it is an oversimplification resulted from the short span (i.e., zero), and a larger span described later will produce more representative results.

Figure 3, discussed in the following paragraph is the expert model for intubation, that will be compared with the model generated by workflow discovery algorithm using each of the datasets, for different values of spans.

6.2. Expert Based Model

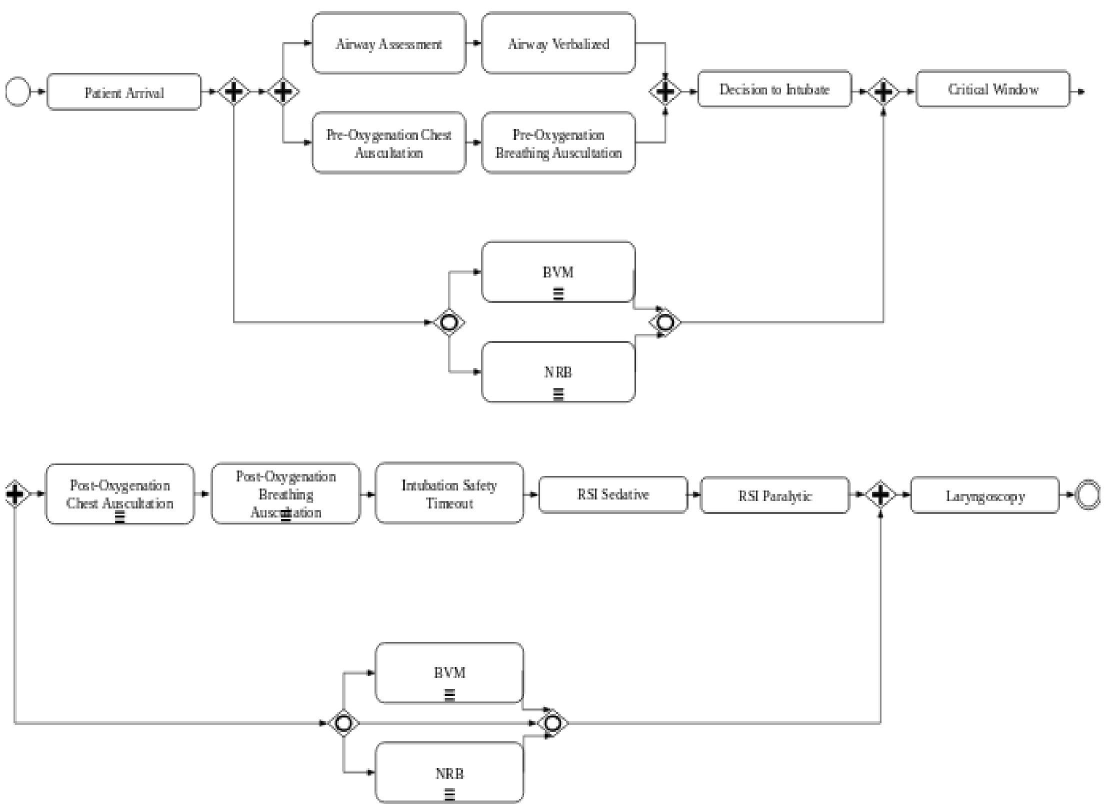


Figure 3 : BPMN diagram for Intubation, derived from expert model

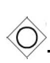
- is the AND split gateway, and - is the OR split gateway

Figure 3 shows a Business Process Modelling Notation (BPMN) representation of the treatment activities in the intubation process. AND split gateway in the diagram means that all the activities linked to the arrow going out from the AND gateway symbol, should occur and could be carried out in parallel. OR split gateway in the diagram mean that any of the paths going out from the OR gateway symbol could be taken. For instance, after 'Patient Arrival', both the paths leading to 'Airway Assessment' and 'Pre-Oxygenation Chest Auscultation' will be taken (also can be executed parallelly) and the path going to 'BVM' (Bag valve mask) and 'NRB' (non-rebreather mask) will also be taken (could be in parallel to 'Airway Assessment' and 'Pre-Oxygenation Chest Auscultation'), but since we have the OR split gateway between 'BVM' and 'NRB', only one of these treatment activities (either BVM or NRB) will be carried out.

Comparing the expert model with the model generated for span of 0 consecutive consensus activity, we see that after patient arrival, 'Airway Assessment' is an important activity that does not appear in the workflow generated by the algorithm. Similarly, there are many activities after 'Critical Window', in figure 3 that are missing from the workflow model in figure 2, thus making it deficient of crucial steps in treatment procedure.

6.3. Comparison Method Used : Levenshtein Distance

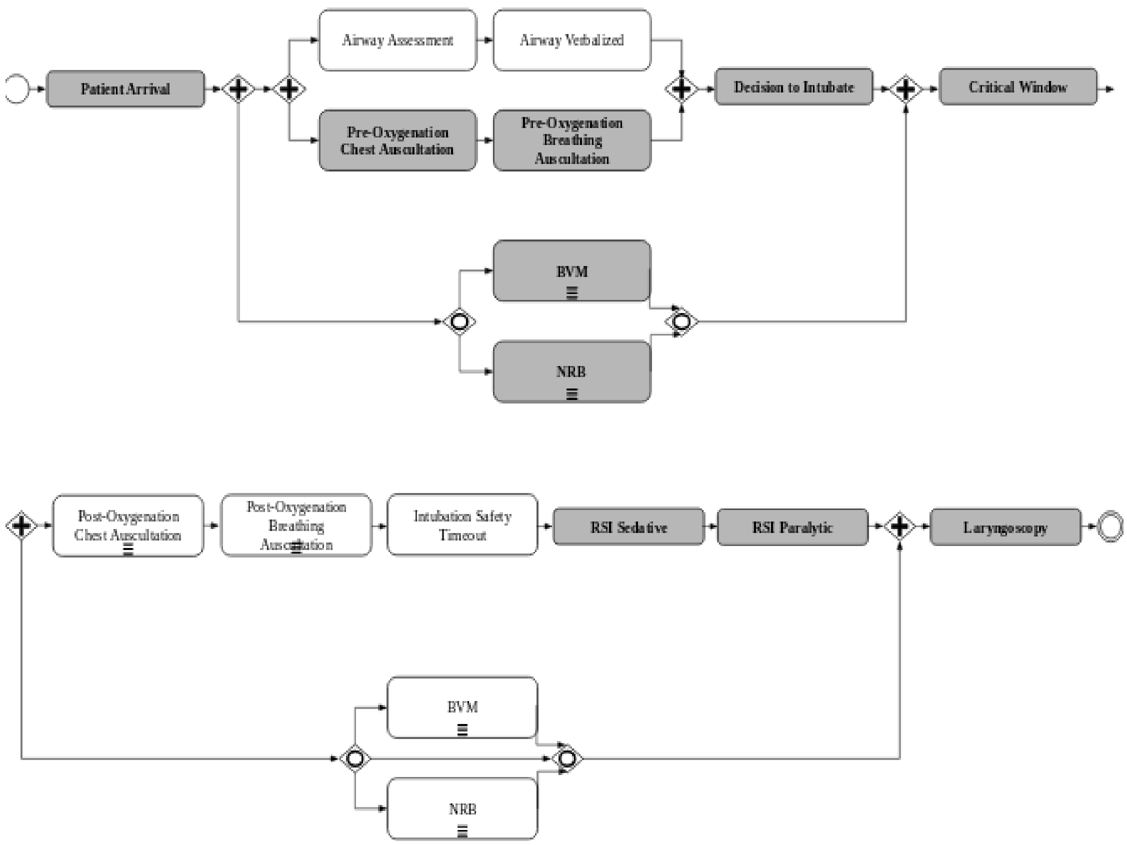


Figure 4 : BPMN diagram for Intubation highlighting the consensus activities for smaller dataset

An illustration of the activities that appeared in the workflow model for span of 0 consecutive consensus activity, and the activities that didn't, is shown in figure 4. The highlighted activities in figure 4 corresponds to the activities that are consensus activities (backbone of the workflow model shown in figure 2).

Figure 4 shows the visual contrast between the consensus activities (greyed out) and the non-consensus activities, but for a better quantitative analysis of the obtained results, we apply the Levenshtein distance [7], to determine the “distance” between the workflow model and the expert-based model. The “distance” mentioned here is calculated as the number of deletions, insertions and substitutions required to transform the workflow model into the expert-based model. We assume that the three operations (deletions, insertions and substitutions) that contribute to the Levenshtein distance are of equal cost (i.e., 1-unit cost operation for each deletions, insertions and substitutions). The activities that are missing from the workflow are counted under insertions required, the activities that are incorrectly added (noise) to the workflow are counted under deletions required, and the activities that required to be substituted with other activities or require any modification in their order of occurrence in the workflow (substitution of their preceding or succeeding activities with other activities in the workflow) are counted under substitutions required. Altogether, the sum of these three required operations (deletions, insertions and substitutions) gives the Levenshtein distance. To illustrate how we calculate the Levenshtein distance, we see that for span of 0 consecutive consensus activity, the workflow model needs insertion of 'Airway Assessment', 'Airway Verbalised', 'Post Oxygenation Chest Auscultation', 'Post Oxygenation Breathing Auscultation', 'Intubation Safety Timeout' and either of 'BVM' or 'NRB'. This means that are 6 insertions required, which makes the Levenshtein distance equal to 6-unit operations. Also, we see that there have been 2 occurrences of 'BVM' and 1 occurrence of 'NRB' before the 'Critical Window', which are redundant because the expert-based model shows that either of the 'BVM' or 'NRB' takes place before the 'Critical Window'. Since this redundancy does not imply an incorrect sequence of occurrence of the treatment activities, we do not classify any of these activities under 'deletions required' or 'substitutions required' while calculating the Levenshtein distance. Lesser the Levenshtein distance, means the "distance" between the workflow model and the expert-based model is lesser and thus more similarity between the workflow and the established "ground truth". We calculate the Levenshtein distance for each of the span values and determine the optimum value at which the Levenshtein distance is least and the workflow model exhibits most precise flow of treatment activities.

6.4. Analysis Using Smaller Intubation Dataset for Span of 1 consecutive consensus activity

Continuing our analysis for Span of 1 consecutive consensus activity, it is clear that the backbone stays the same, and we might get additional common but dispersed activities, if there are non-consensus activities spread across a span of 1 consensus activity, that has combined probability of occurrence more than the threshold (0.37). For the span of 1 consecutive consensus activity we get the model shown in figure 5, as follows -

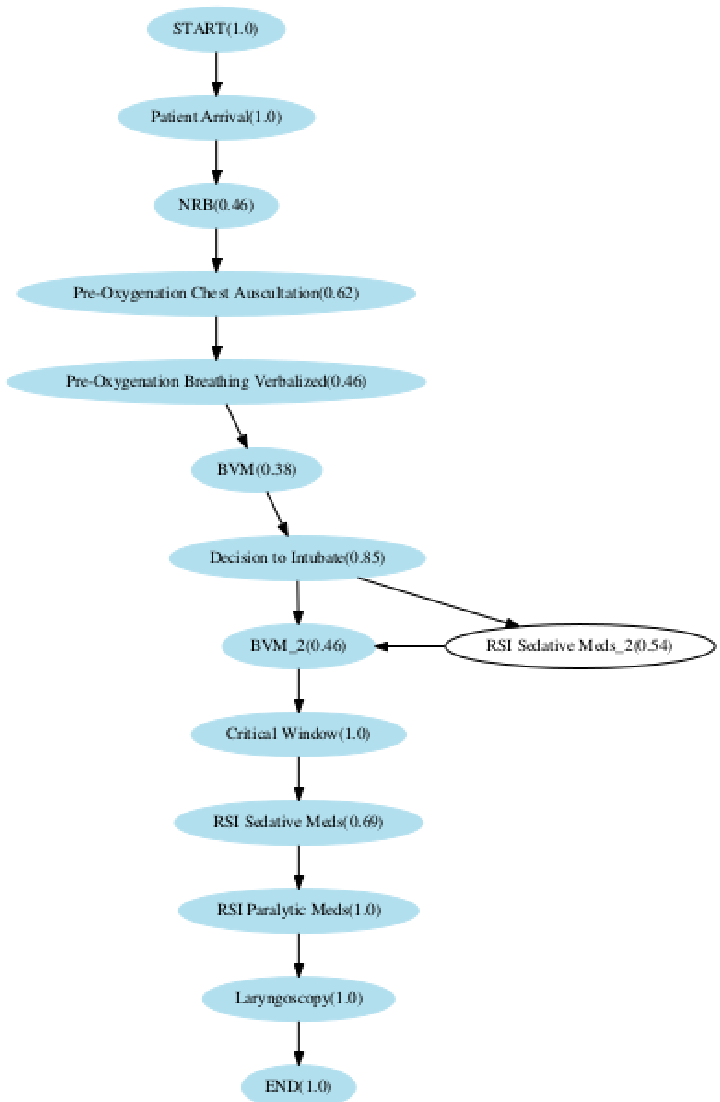


Figure 5 : Workflow model for smaller dataset at span value of 1 consecutive consensus activity

Figure 5 shows the workflow model for span of 1 consecutive consensus activity that has a branch, unlike the linear workflow model shown in figure 2 for the span of 0 consecutive consensus activity that was an oversimplification resulted from the short span (i.e., zero). From figure 5, we get ‘RSI Sedative Meds_2’ (where RSI stands for rapid sequence induction) as the common but dispersed activity. However, if we compare the model with the expert model in figure 3, we can infer that ‘RSI Sedative Meds_2’ is an activity that should appear after the ‘Critical Window’, but in figure 5, it appears to occur before ‘Critical Window’. Additionally, since we already have the activity ‘RSI Sedative Meds’ in the backbone that appears after ‘Critical Window’, and thus seems more appropriate due to its correct order of appearance, we can classify ‘RSI Sedative Meds_2’ as noise. Thus, we can conclude that for the span of 1 consecutive consensus activity, we don't have an improvement over the span of 0 consecutive consensus activity - the model stays similar i.e. deficient of crucial treatment activities, with an added noise, i.e., ‘RSI Sedative Meds_2’. Since, ‘RSI Sedative Meds_2’ is the noise, we can count it as the required deletion while calculating the Levenshtein distance. In addition to the 1 deletion, the workflow model still needs the insertion of ‘Airway Assessment’, ‘Airway Verbalised’, ‘Post Oxygenation Chest Auscultation’, ‘Post Oxygenation Breathing Auscultation’, ‘Intubation Safety Timeout’ and either of ‘BVM’ or ‘NRB’ (as calculated for Levenshtein distance of the span of 0 consecutive consensus activity). This means that are 6 insertions and 1 deletion required, which makes the Levenshtein distance equal to 7-unit operations.

6.5. Analysis Using Smaller Intubation Dataset for Span of 2 consecutive consensus activities

Continuing our analysis for Span of 2 consecutive consensus activities, we get the following model in figure 6 -

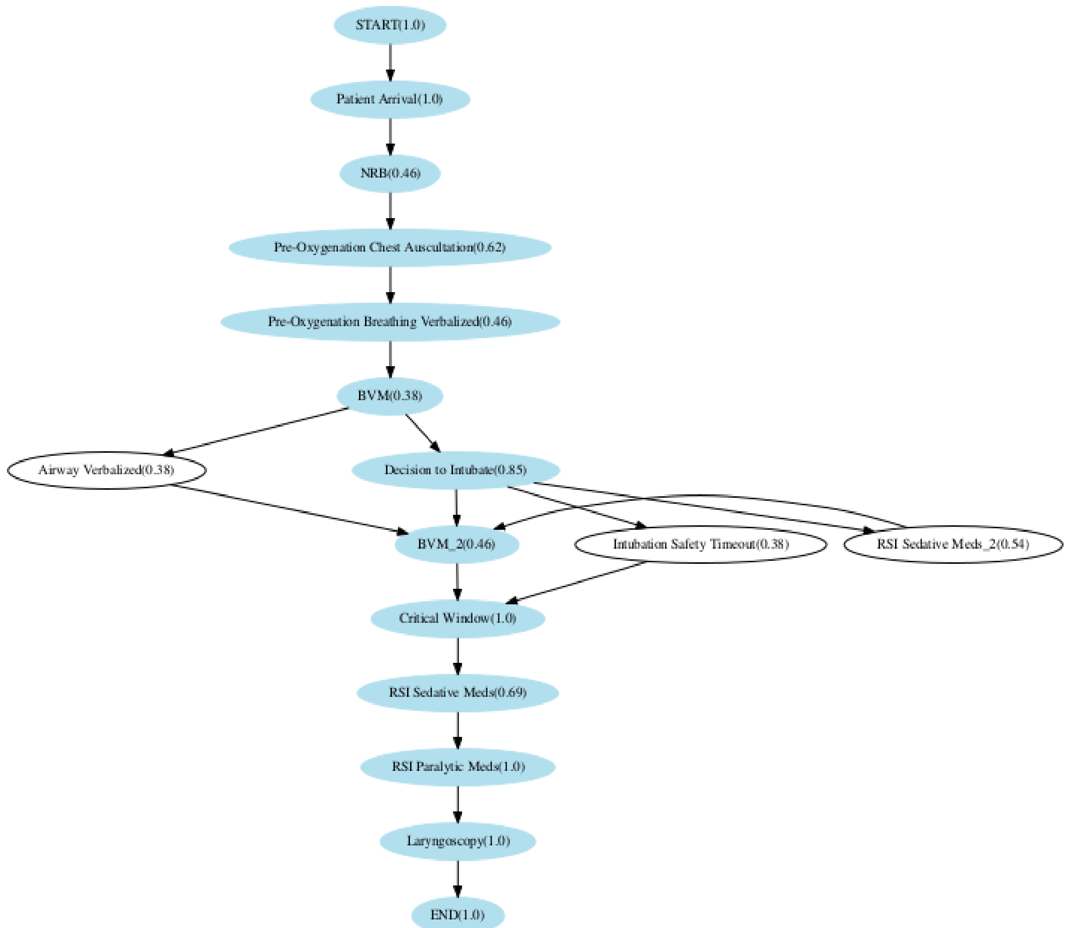


Figure 6 : Workflow model for smaller dataset at span of 2 consecutive consensus activities

In figure 6, it is evident that the model has more branches than we had for span of 1 consecutive consensus activity. These branches corresponds to more common but dispersed activities introduced in the workflow for the span of 2 consecutive consensus activities, than we had for span of 1 consecutive consensus activity, but with these additional activities, we need to compare the workflow model with the expert model (figure 3) to check if these are relevant to the procedure or are just added noise. Comparing the expert model with the model generated for span of 2 consecutive consensus activities, we see that the ‘Airway Verbalized’ appears without the prior occurrence of ‘Airway Assessment’, but it should have been after the appearance of ‘Airway Assessment’ and similarly, ‘Intubation Safety Timeout’ appears before ‘Critical Window’, but it should have been after the ‘Critical Window’. Therefore, we can infer that ‘Airway Verbalized’ and ‘Intubation Safety Timeout’, are out of order in the workflow, and thus doesn't add to the relevancy of the model. At this point we have additional noise of ‘Airway Verbalized’ and ‘Intubation Safety Timeout’ as compared to span of 1 consecutive consensus activity, and the generated model isn't very informative as it still lacks about 5 of the treatment activities, when compared to the figure 3. Since, ‘RSI Sedative Meds_2’, ‘Airway Verbalized’ and ‘Intubation Safety Timeout’ are the noise, we can count it as the required deletion while calculating the Levenshtein distance. In addition to the 3 deletions, the workflow model still needs the insertion of ‘Airway Assessment’, ‘Airway Verbalised’, ‘Post Oxygenation Chest Auscultation’, ‘Post Oxygenation Breathing Auscultation’ and either of ‘BVM’ or ‘NRB’ (as calculated for Levenshtein distance of the span of 0 consecutive consensus activity). This means that are 5 insertions and 3 deletions required, which makes the Levenshtein distance equal to 8-unit operations.

6.6. Analysis Using Smaller Intubation Dataset for Span of 3 consecutive consensus activities

Taking our analysis further for the span value of 3 consecutive consensus activities, we get the following model in figure 7 -

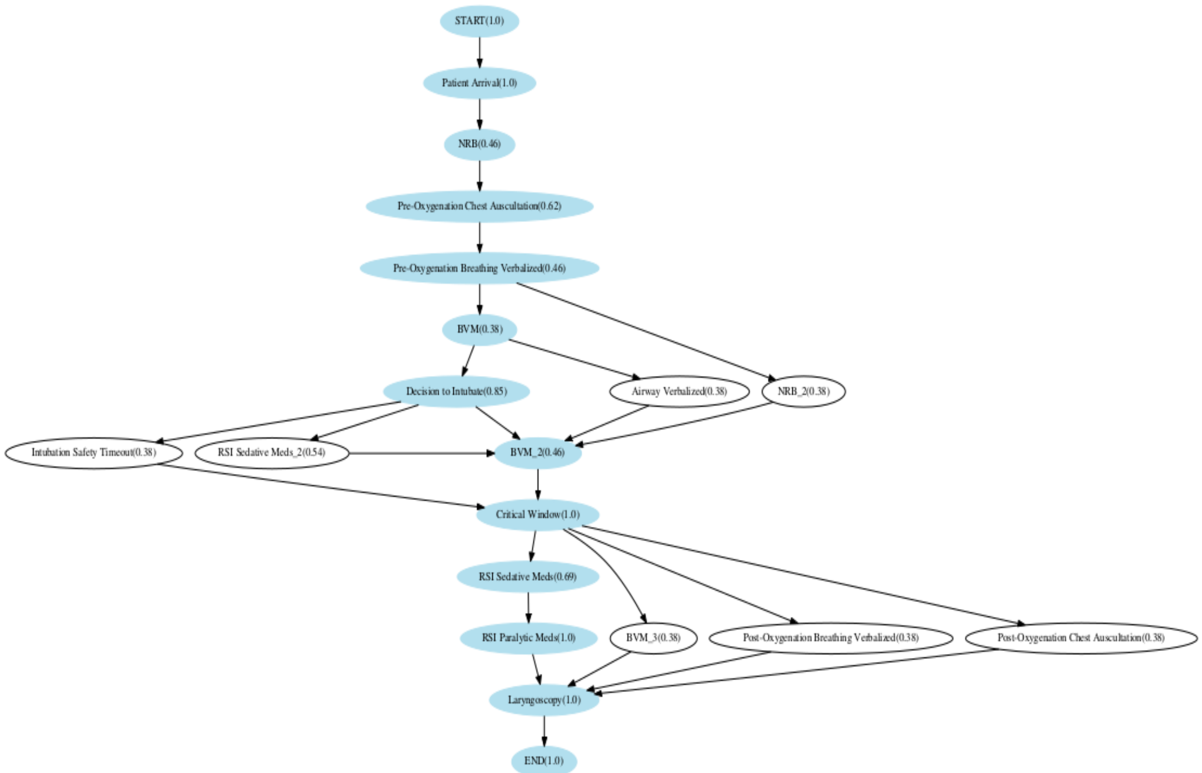


Figure 7 : Workflow model for smaller dataset at span of 3 consecutive consensus activities

From figure 7, it is clear that we have additional branches, that corresponds to more common but dispersed activities as compared to the model generated for span of 2 consecutive consensus activities. We get NRB, BVM, Post-Oxygenation Breathing Verbalized and Post-oxygenation Chest Auscultation in addition to the workflow model generated for span of 2 consecutive consensus activities. The addition of BVM, Post-Oxygenation Breathing Verbalized and Post-oxygenation Chest Auscultation adds up to the relevancy of workflow information and could be classified as a correct addition of common but dispersed activities to form an accurate and interpretable workflow. The addition of NRB could be seen as a trivial addition to the workflow, because the information was already complete, in the sense that we already had NRB in the workflow. On comparison with the expert model, we see that minor details such as lack of ‘Airway Assessment’ activity are still prevalent in the workflow, and the noise remains the same as previous models (3 activities : ‘Airway Verbalized’, ‘RSI_Sedative Meds_2’ and ‘Intubation Safety Timeout’). However, the generated workflow consists of almost all of the treatment activities for intubation and is closer in procedural order of the activities as compared to other models generated for Span values ranging from 0-2 consecutive consensus activities. The noise stays the same as for span of 2 consecutive consensus activities, but with the advantage of increased accuracy (with the added correct order of occurrence of BVM, Post-Oxygenation Breathing Verbalized and Post-oxygenation Chest Auscultation), span value of 3 consecutive consensus activities is better of all the span values seen before. Since, ‘RSI Sedative Meds_2’, ‘Airway Verbalized’ and ‘Intubation Safety Timeout’ are the noise, we can count it as the required deletion while calculating the Levenshtein distance. In addition to the 3 deletions, the workflow model still needs the insertion of ‘Airway Assessment’ and ‘Airway Verbalised’. This means that are 2 insertions and 3 deletions required, which makes the Levenshtein distance equal to 5-unit operations.

6.7. Analysis Using Smaller Intubation Dataset for Span of 4 consecutive consensus activities

Continuing our experimental analysis further, we test the model generated for span of 4 consecutive consensus activities, and the generated workflow is shown below in figure 8 -

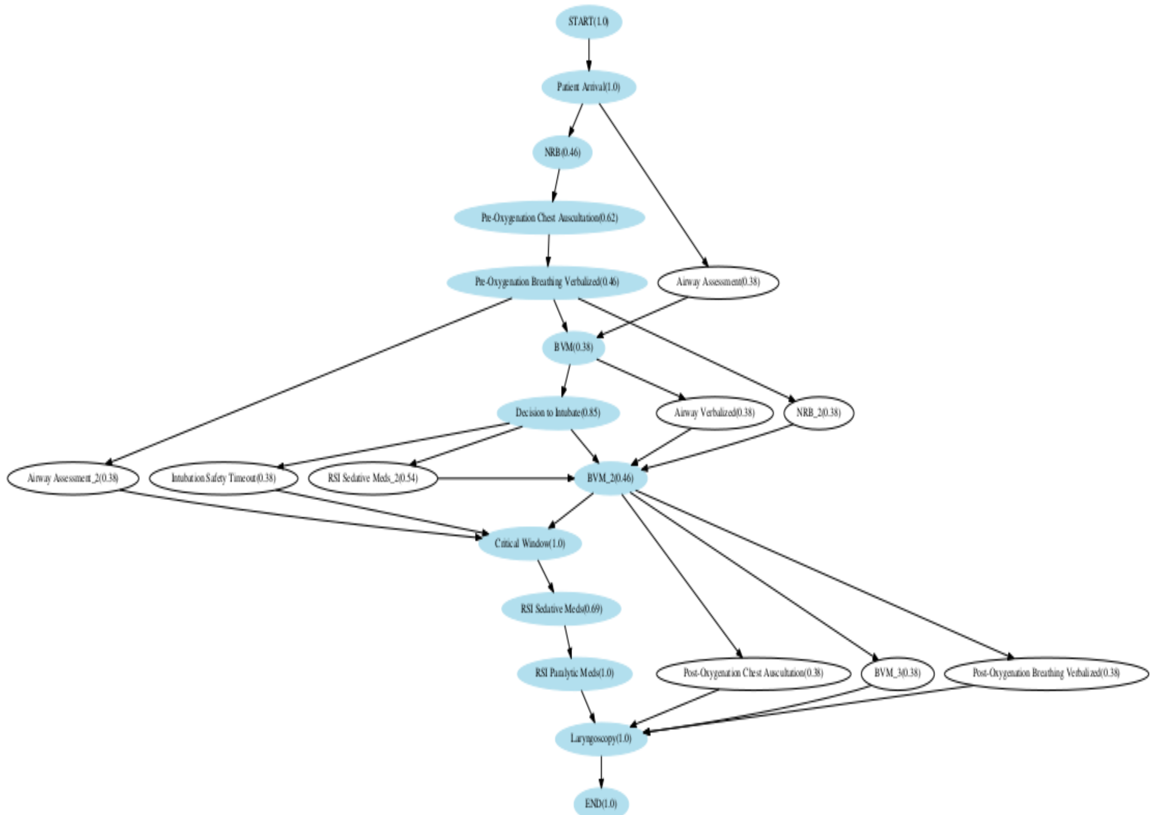


Figure 8 : Workflow model for smaller dataset at span of 4 consecutive consensus activities

The generated workflow model for span of 4 consecutive consensus activities is somewhat similar to what we got for span of 3 consecutive consensus activities, but there is additional ‘Airway Assessment’ between ‘Patient Arrival’ and ‘BVM’, in the workflow that fulfills the shortcomings in terms of information that previous model was lacking. Figure 7 lacks ‘Airway Assessment’ in the workflow, which is now introduced in figure 8 for span of 4 consecutive consensus activities. Comparing the model shown in figure 8 with the expert model in figure 3, we see that ‘Airway Assessment’ is in order with the flow of the generated model (appears correctly between the occurrence of ‘Patient Arrival’ and ‘BVM’ before the ‘Decision to Intubate’), and couldn't be classified as noise, however there is another ‘Airway Assessment’ between ‘Pre-oxygenated Breathing Verbalised’ and ‘Critical Window’, that appears to be a trivial addition to the workflow, since we already have an ‘Airway Assessment’ that appears between ‘Patient Arrival’ and ‘BVM’ as it should be (when compared to the expert model in figure 3). Additionally, we also lost the preciseness in the order of BVM, Post-Oxygenation Chest Auscultation and Post-Oxygenation Breathing Verbalised, that now appears between another ‘BVM’, and ‘Laryngoscopy’, instead of more apt representation in the previous model where these activities existed between ‘Critical Window’ and ‘Laryngoscopy’. Since, ‘RSI Sedative Meds_2’, ‘Airway Verbalized’ and ‘Intubation Safety Timeout’ are the noise, we can count it as the required deletion while calculating the Levenshtein distance. Also, the second ‘Airway Assessment’ that appears between ‘Pre-oxygenated Breathing Verbalised’ and ‘Critical Window’, is a required deletion, since we have established that it is an inaccurate and trivial addition to the workflow. In addition to the 4 deletions, the workflow model still needs an insertion of ‘Airway Verbalised’. Along with insertions and deletions, we address the lost accuracy in the order of occurrence of BVM, Post-Oxygenation Chest Auscultation and Post-Oxygenation Breathing Verbalised, (appears between another ‘BVM’, and ‘Laryngoscopy’, instead of more apt representation in the previous model where these activities existed between ‘Critical Window’ and ‘Laryngoscopy’), and count them as the required substitutions. This means that there is 1 insertion, 4 deletions and 3 substitutions required, which makes the Levenshtein distance equal to 8-unit operations. It is noteworthy at this point that span of 4 consecutive consensus activities generates a workflow model that has added noise and lesser accuracy as compared to the previous model generated for span value of 3 consecutive consensus activities - implying that span value of 3 consecutive consensus activities, is the optimum value of span that we have seen till now.

6.8. Analysis Using Smaller Intubation Dataset for Span of 5 consecutive consensus activities

Taking our analysis further to span of 5 consecutive consensus activities, we get the workflow model shown below in figure 9 -

Figure 9 : Workflow model for smaller dataset at span of 5 consecutive consensus activities

The workflow model generated for span of 5 consecutive consensus activities is exactly the same as the model for span of 4 consecutive consensus activities. There is no additional common but dispersed activity in the workflow. Our analysis remains the same for span of 5 consecutive consensus activities as it was for span of 4 consecutive consensus activities, and our Levenshtein distance equals to 8-unit operations.

<u>6.9. Analysis Using Smaller Intubation Dataset for Span of 6 consecutive consensus activities</u>

Continuing to the span of 6 consecutive consensus activities, we get the following model in figure 10 -

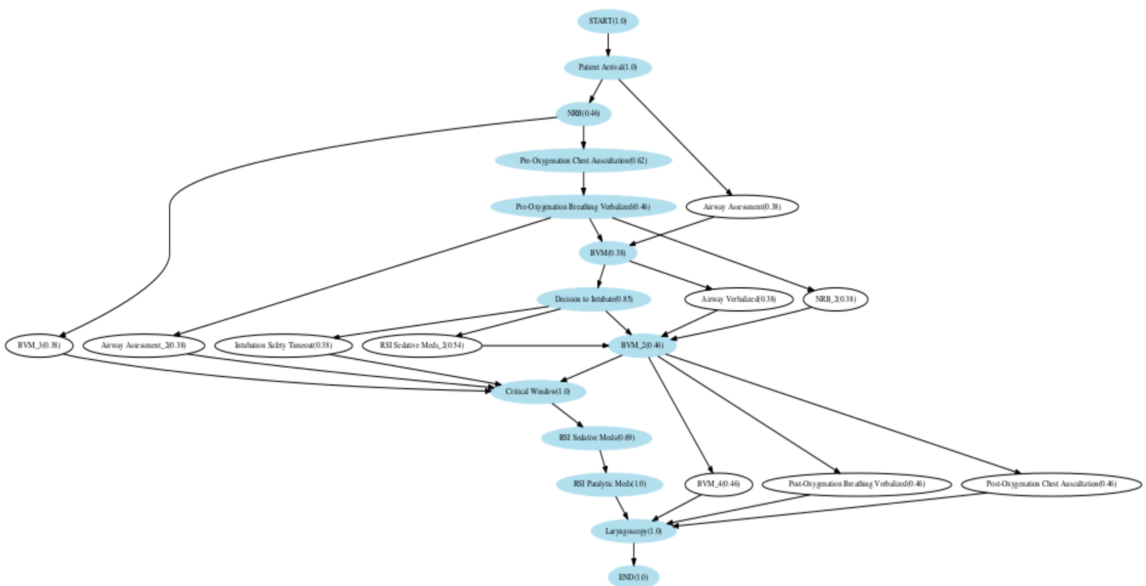


Figure 10 : Workflow model for smaller dataset at span of 6 consecutive consensus activities

At the span of 6 consecutive consensus activities, all the activities and the generated workflow model remains the same, except for an additional BVM between NRB and Critical Window. On comparison with the expert model, it can be said that this addition to the workflow wasn't necessary, because we already have a ‘BVM’ in the backbone that corresponds to the BVM before critical window. However, this addition isn't incorrect or out of order, to be classified under the category of noise, it seems to be just a redundant addition. Since this redundancy does not imply an incorrect sequence of occurrence of the treatment activity, we do not classify BVM under ‘deletions required’ or ‘substitutions required’ while calculating the Levenshtein distance. Thus, the Levenshtein distance is still equal to 8-unit operations. It can be inferred that for span of 6 consecutive consensus activities, it isn't an improvement over the established notion that the span of 3 consecutive consensus activities is the optimum value till now as we iterate over different values of span.

6.10. Analysis Using Smaller Intubation Dataset for Span of 7 consecutive consensus activities

We get the following model in figure 11, for a span value of 7 consecutive consensus activities -

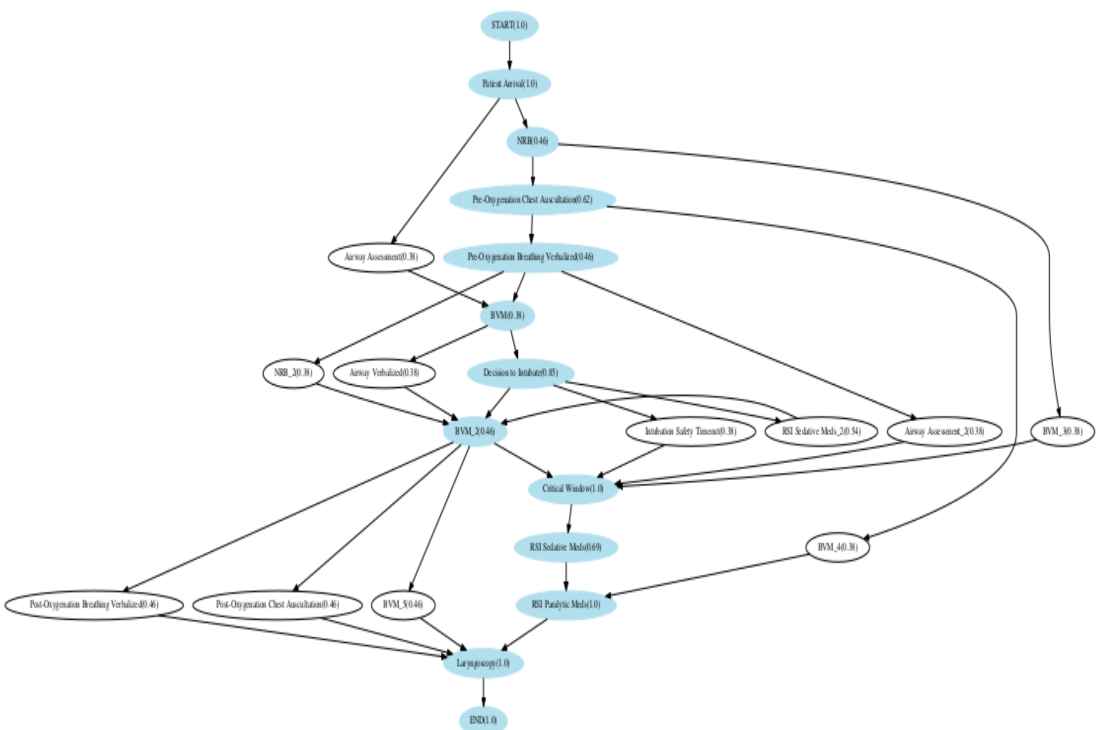


Figure 11 : Workflow model for smaller dataset at span of 7 consecutive consensus activities

There is no significant difference in the model generated for the span value of 7 consecutive consensus activities and the model generated for span value of 6 consecutive consensus activities, except for an additional ‘BVM’ between ‘Pre Oxygenation Chest Auscultation’ and ‘RSI Paralytic Meds’, on comparing the model with expert model, the additional BVM appears to be a noise because procedurally there is no such occurrence of BVM, that occurs between ‘Pre Oxygenation Chest Auscultation’ and ‘RSI Paralytic Meds’. This addition of BVM thus, does not contribute to the accuracy of information carried by the model, only increasing the complexity slightly. Since, ‘BVM’, ‘RSI Sedative Meds_2’, ‘Airway Assessment’, ‘Airway Verbalized’ and ‘Intubation Safety Timeout’ are the noise, we can count it as the required deletion while calculating the Levenshtein distance. In addition to the 5 deletions, the workflow model still needs a correct insertion of ‘Airway Verbalised’. Along with insertions and deletions, we have required substitutions (BVM, Post-Oxygenation Chest Auscultation and Post-Oxygenation Breathing Verbalised) from the previous discussion for span of 4 consecutive consensus activities. This means that there is 1 insertion, 5 deletions and 3 substitutions required, which makes the Levenshtein distance equal to 9-unit operations. It is noteworthy at this point that span of 4 consecutive consensus activities generates a workflow model that has added noise and lesser accuracy as compared to the previous model generated for span value of 3 consecutive consensus activities - implying that span value of 3 consecutive consensus activities, is the optimum value of span that we have seen till now.

6.11. Analysis Using Smaller Intubation Dataset for Span of 8 consecutive consensus activities

Moving further to span of 8 consecutive consensus activities, we get the diagram as shown in figure 12 below -

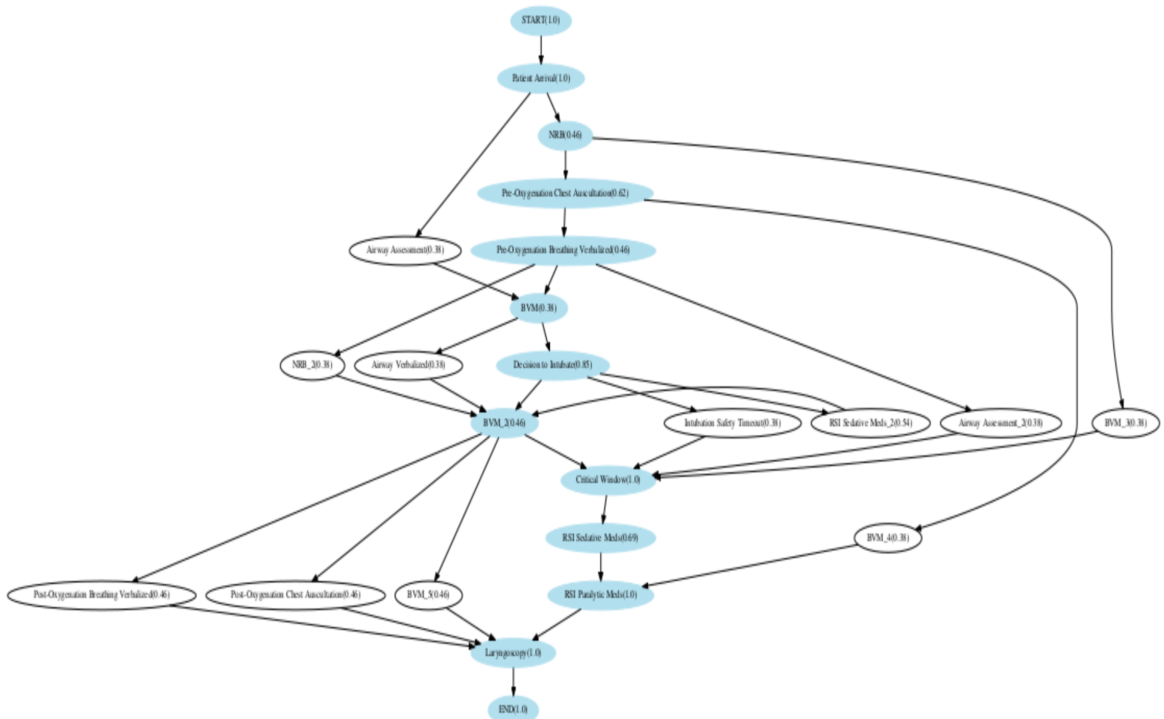


Figure 12 : Workflow model for smaller dataset at span of 8 consecutive consensus activities

The model generated for the span value of 8 consecutive consensus activities shown in figure 12, is the same as what we got for span of 7 consecutive consensus activities. Hence, our analysis remains the same as it was for span of 7 consecutive consensus activities, and the Levenshtein distance is equal to 9-unit operations.

<u>6.12. Analysis Using Smaller Intubation Dataset for Span of 9 consecutive consensus activities</u>

Moving over to our analysis for the span of 9 consecutive consensus activities, which is shown below in figure 13 -

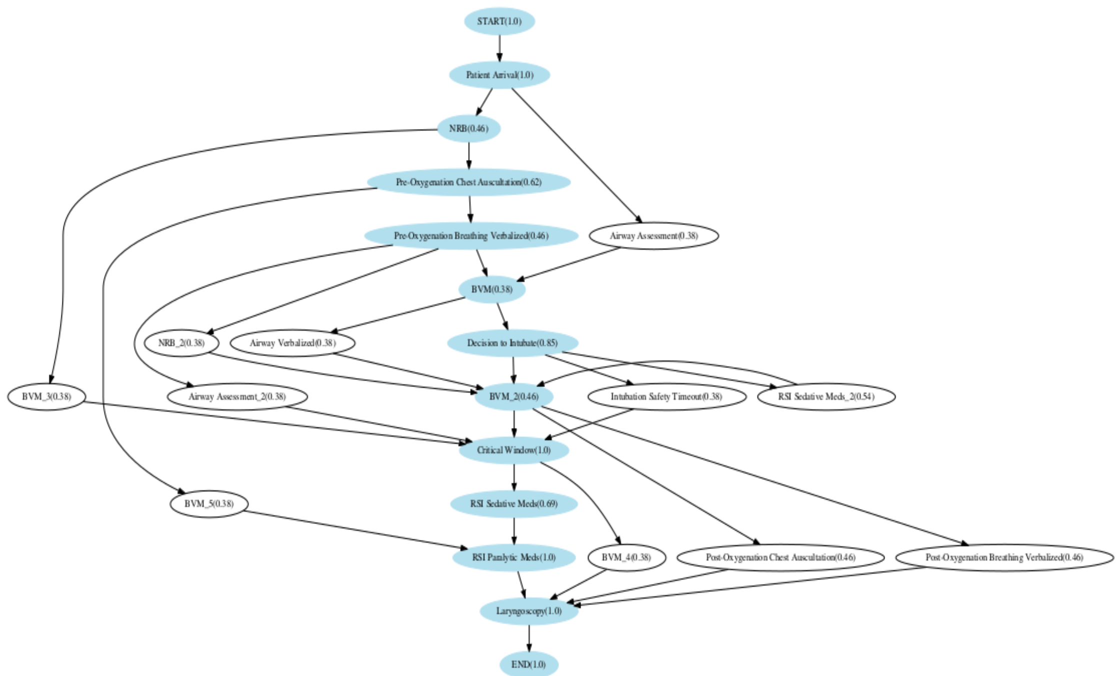


Figure 13 : Workflow model for smaller dataset at span of 9 consecutive consensus activities

The workflow model generated for span of 9 consecutive consensus activities exhibits similarities with the model generated for span of 8 consecutive consensus activities, but with the exception of BVM being reordered to exist between critical window and laryngoscopy, instead of existing between another BVM and laryngoscopy. On comparing with the expert model in figure 3, it appears that this order is rather more appropriate, eliminating the out of order occurrence of BVM for span of 8 consecutive consensus activities, and having it occur in-order for span of 9 consecutive consensus activities. Thus, we see that we lost the 'BVM' that was added as noise in the previous workflows and have another BVM added to the workflow shown in figure 13, that is correct. For figure 13, we have 'BVM', 'RSI Sedative Meds_2', 'Airway Assessment', ‘Airway Verbalized’ and ‘Intubation Safety Timeout’ as the noise, and therefore we can count these as the required deletion while calculating the Levenshtein distance. In addition to the 5 deletions, the workflow model still needs a correct insertion of ‘Airway Verbalised’. Along with insertions and deletions, we have required substitutions (Post-Oxygenation Chest Auscultation and Post-Oxygenation Breathing Verbalised) from the previous discussion for span of 4 consecutive consensus activities. This means that there is 1 insertion, 5 deletions and 2 substitutions required, which makes the Levenshtein distance equal to 8-unit operations. At this value, we see the model poses slightly better flow, but still with additional noise when compared to the model generated for span of 3 consecutive consensus activities, couldn't be seen as an improvement over the workflow model for the span of 3 consecutive consensus activities. With the disadvantage of increased noise and complexity (by trivial additions), we can still consider span of 3 consecutive consensus activities to be a better choice over span of 9 consecutive consensus activities.

6.13. Analysis Using Smaller Intubation Dataset for Span of 10 consecutive consensus activities

For span of 10 consecutive consensus activities, we get the workflow as shown in figure 14 below -

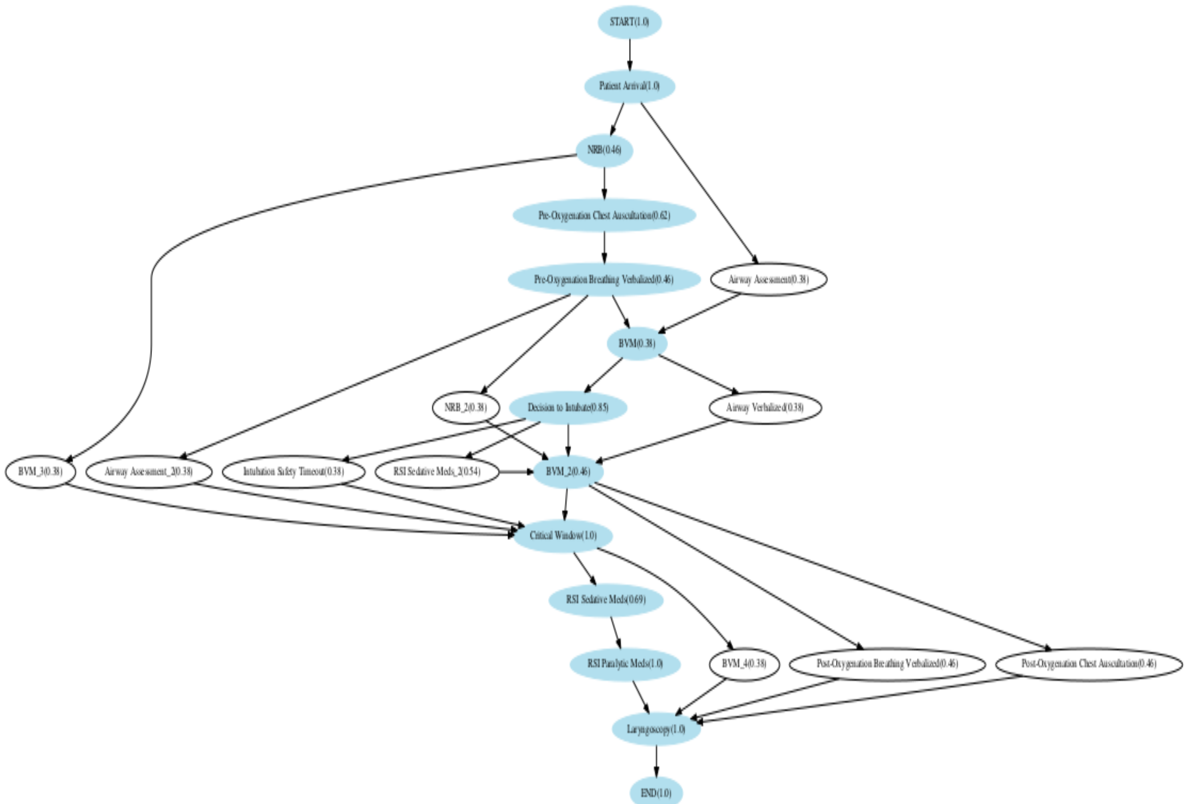


Figure 14 : Workflow model for smaller dataset at span of 10 consecutive consensus activities

Span of 10 consecutive consensus activities eliminates the noise in the model for span of 9 consecutive consensus activities, by getting rid of the ‘BVM’ between ‘Pre-oxygenation Chest Auscultation’ and ‘RSI Paralytic Meds’. The workflow model for span of 10 consecutive consensus activities is the same as for span of 9 consecutive consensus activities but with a decreased noise (of ‘BVM’ between ‘Pre-oxygenation Chest Auscultation’ and ‘RSI Paralytic Meds’ that existed in the previous workflow model), thus making it better than span of 9 consecutive consensus activities, but still poses more noise over span of 3 consecutive consensus activities. We still have ‘RSI Sedative Meds_2’, ‘Airway Assessment’, ‘Airway Verbalized’ and ‘Intubation Safety

Timeout' as the noise, and therefore we can count these as the required deletion while calculating the Levenshtein distance. In addition to the 4 deletions, the workflow model still needs a correct insertion of 'Airway Verbalised'. Along with insertions and deletions, we have required substitutions (Post-Oxygenation Chest Auscultation and Post-Oxygenation Breathing Verbalised) from the previous discussion for span of 4 consecutive consensus activities. This means that there is 1 insertion, 4 deletions and 2 substitutions required, which makes the Levenshtein distance equal to 7-unit operations.

6.14. Analysis Using Smaller Intubation Dataset for Span of 11 consecutive consensus activities

Moving ahead for span of 11 consecutive consensus activities, we get the workflow -

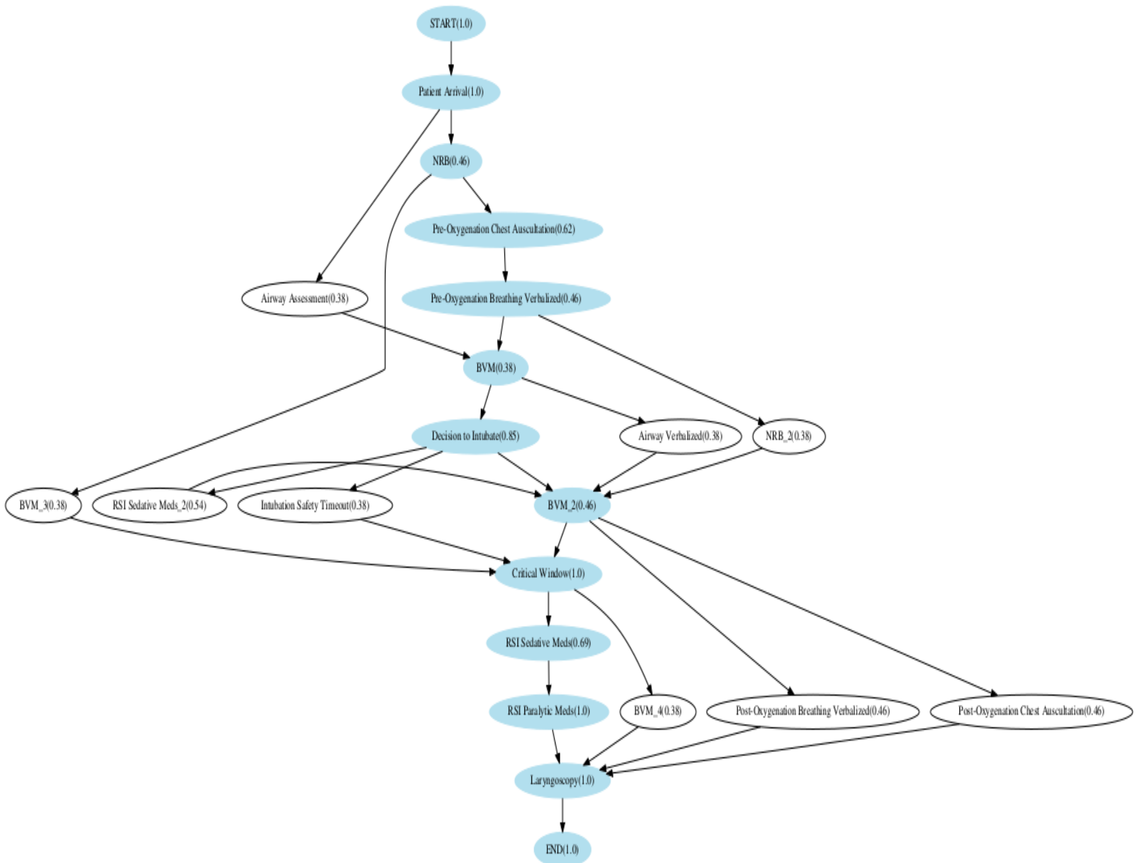


Figure 15 : Workflow model for smaller dataset at span of 11 consecutive consensus activities

The workflow model generated for span of 11 consecutive consensus activities, further eliminates the trivial addition of 'Airway Assessment' between 'Pre-oxygenated Breathing Verbalised' and 'Critical Window', that was added at span of 4 consecutive consensus activities, and existed in all of the models till the span of 10 consecutive consensus activities. Thus, span of 11 consecutive consensus activities could be seen as an improvement over the span of 10 consecutive consensus activities, but still stand noisier and slightly more complex (by the addition of a trivial activity) than span of 3 consecutive consensus activities.

We still have ‘RSI Sedative Meds_2’, ‘Airway Verbalized’ and ‘Intubation Safety Timeout’ as the noise, and therefore we can count these as the required deletion while calculating the Levenshtein distance. In addition to the 3 deletions, the workflow model still needs a correct insertion of ‘Airway Verbalised’. Along with insertions and deletions, we have required substitutions (Post-Oxygenation Chest Auscultation and Post-Oxygenation Breathing Verbalised) from the previous discussion for span of 4 consecutive consensus activities. This means that there is 1 insertion, 3 deletions and 2 substitutions required, which makes the Levenshtein distance equal to 6-unit operations.

Given that we have 11 consensus activities (excluding start and end), forming the backbone of our workflow model, our span analysis ranges only from 0 to 11 consecutive consensus activities. The range of span varies between 0 and the number of consensus activities we get for a particular case. Thus, if we go for span values beyond 11 consecutive consensus activities, the model remains the same, taking into account the span of 11 consecutive consensus activities. For the span value of 12 consecutive consensus activities, we get the model shown below in figure 16 that is exactly the same model as we got for span of 11 consecutive consensus activities, hence the analysis remains the same -

Figure 16 : Workflow model for smaller dataset at span of 12 consecutive consensus activities

7. Results for Smaller Intubation Dataset

Thus, for the given dataset, we have seen the varying accuracy, noise and Levenshtein distance in the generated workflow model, at different values of span, ranging from span values of 0 to 11 consecutive consensus activities. The Levenshtein distance was least for the workflow model of span of 3 consecutive consensus activities. We can conclude that, for this particular case, Span of 3 consecutive consensus activities turns out to be the optimum value of span, because for span of 3 consecutive consensus activities we get a model closest to the expert model, with the maximum number of correct activities, least noise and least complexity. While considering the permissible amount of noise, we also need to look at the accuracy of workflow. We may get a workflow model with less information and less noise, and another one with more information and slightly more amount of noise than the other model - in that case the judgement lies in the fact that we need to have more informational accuracy within an acceptable range of noise. For the models generated at span of 9-11 consecutive consensus activities, we get the same number of correct common but dispersed activities as span of 3 consecutive consensus activities, but at the cost of increased noise and trivial additions to the workflow, thereby increasing complexity. A more clear picture of the discussion could be seen in the condensation of our analysis for this particular case of Intubation procedure, shown in the table 7.1 that compares different values of span, to the changes in accuracy (inferred by Correct activities and Noise column) and complexity (inferred by Trivial activities and Noise) -

Span (of consecutive consensus activities)	**Number of consensus activities**	**Total number of activities**	**Correct activities (common but dispersed)**	**Trivial activities**	**Noise**	**Levenshtein Distance**
0	10	11	0	0	0	6
1	11	12	0	0	1	7
2	11	14	0	0	3	8
3	**11**	**18**	**3**	**1**	**3**	**5**
4	11	20	2	2	5	8
5	11	20	2	2	5	8
6	11	21	2	3	5	8
7	11	22	2	3	6	9
8	11	22	2	3	6	9
9	11	22	3	3	5	8
10	11	21	3	3	4	7
11	11	20	3	2	4	6

Table 7.1 : Workflow composition at different values of span (smaller intubation dataset)

We have seen the model behavior vary for different values of span that consider the interleaved common but dispersed activities between consensus activities. For a real-world application, this analysis of span is important because we cannot afford to skip an integral part of the treatment procedure, based on the set threshold for its probability of occurrence in a data set. Using the span, we iterate over the alignment matrix to locate all such common but dispersed activities that couldn't make it to the workflow in the first pass. We have also simultaneously looked at the noise, to reduce the occurrence of trivial activities in the workflow.

8. Description of Dataset Used for Larger Intubation Dataset

The alignment matrix that we get is subject to the dataset that we use - hence resulting in different optimum values of span for different cases of dataset used. For the case discussed, we used the smaller dataset for Intubation procedure. Going ahead with our analysis, in the following section we analyze the same procedure of Intubation, but with a larger dataset (of 1240 cases).

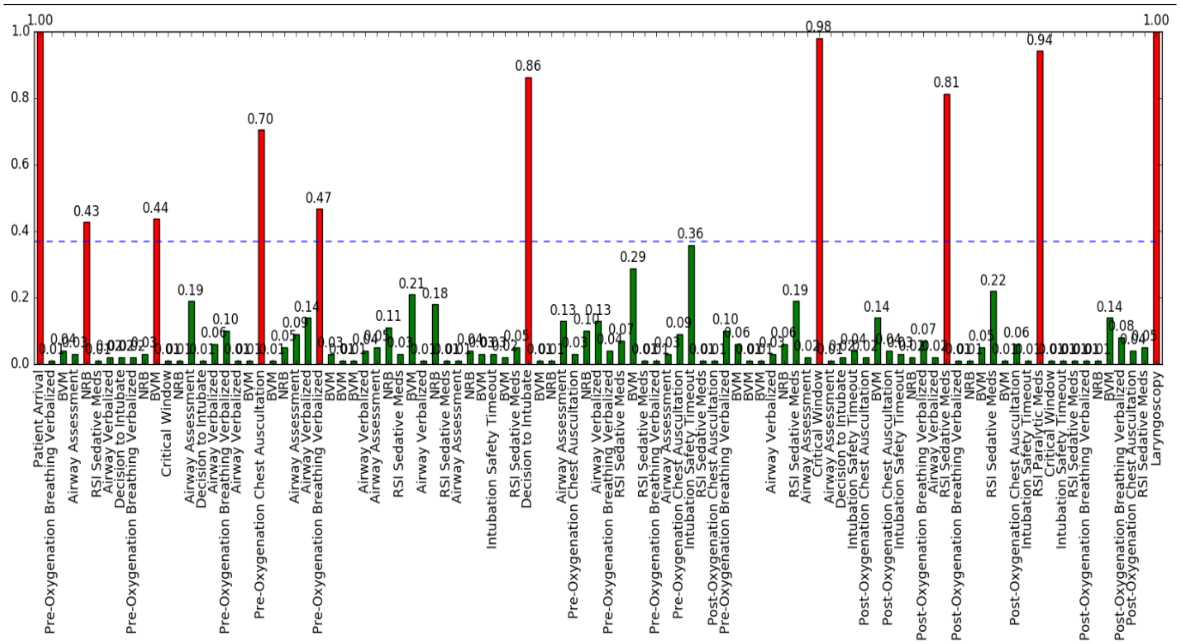


Figure 17 : Treatment activities vs probability of occurrences plot for larger intubation dataset

Figure 17 shows treatment activities for intubation, plotted against their probability of occurrence, for the larger intubation dataset. Similar to figure 1, these activities are ordered by their appearance in the alignment matrix. Every activity trace starts with the activity “Patient Arrival”, and ends with the activity “Laryngoscopy”, therefore these activities don't appear anywhere else in the plot except for the first and last columns. However, other activities might be performed in different orders for different activity traces, and thus, appear in the order of their appearance in alignment matrix. The threshold (□) is set to 0.37, thus classifying all the activities with probability of occurrence greater than 0.37 as consensus activities - red colored that form the backbone of our workflow model, and the rest of them as non-consensus activities - green colored. It is noteworthy, that we have lesser number of consensus activities (10 consensus activities) in this case for larger dataset as compared to the smaller dataset (11 consensus activities). We have highlighted earlier in our discussion that depending upon the dataset we use, we can have different trace alignment matrices, thus giving different sets of consensus and non-consensus activities - as evident in this case, we get different alignment results for the same procedure but with more data. In this case we will have the value of span ranging from 0 to 10 consecutive consensus activities and if we go for span values beyond 10 consecutive consensus activities, the model remains the same, taking into account the span value of 10 consecutive consensus activities.

9. Analysis for Larger Intubation Dataset

9.1. Analysis Using Larger Intubation Dataset for Span of 0 consecutive consensus activity

For Span of 0 consecutive consensus activity, we go through 0 consecutive consensus columns to determine the common but dispersed activities. We would have no common but dispersed activity, and the workflow model will only comprise of the backbone formed by the consensus activities. The diagram shown below in figure 18 displays the workflow model generated for the larger dataset, at span of 0 consecutive consensus activity (shows the consensus activities, i.e. the backbone).

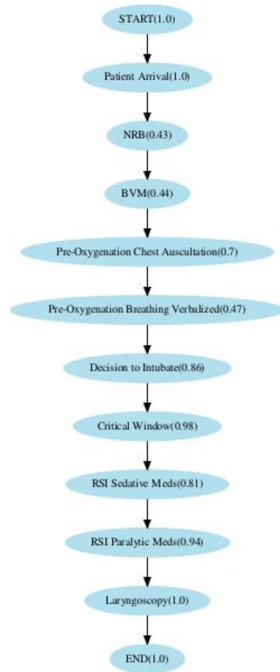


Figure 18 : Workflow model for larger dataset at span of 0 consecutive consensus activity

The workflow model generated for the span of 0 consecutive consensus activity, is shown in figure 18, where we see that the model contains only the backbone structure with no common but dispersed activity. Figure 18 shows the linear workflow model, with no branches because it is an oversimplification resulted from the short span (i.e., zero), and a larger span described later will produce more representative results.

9.2. Comparison Method Used

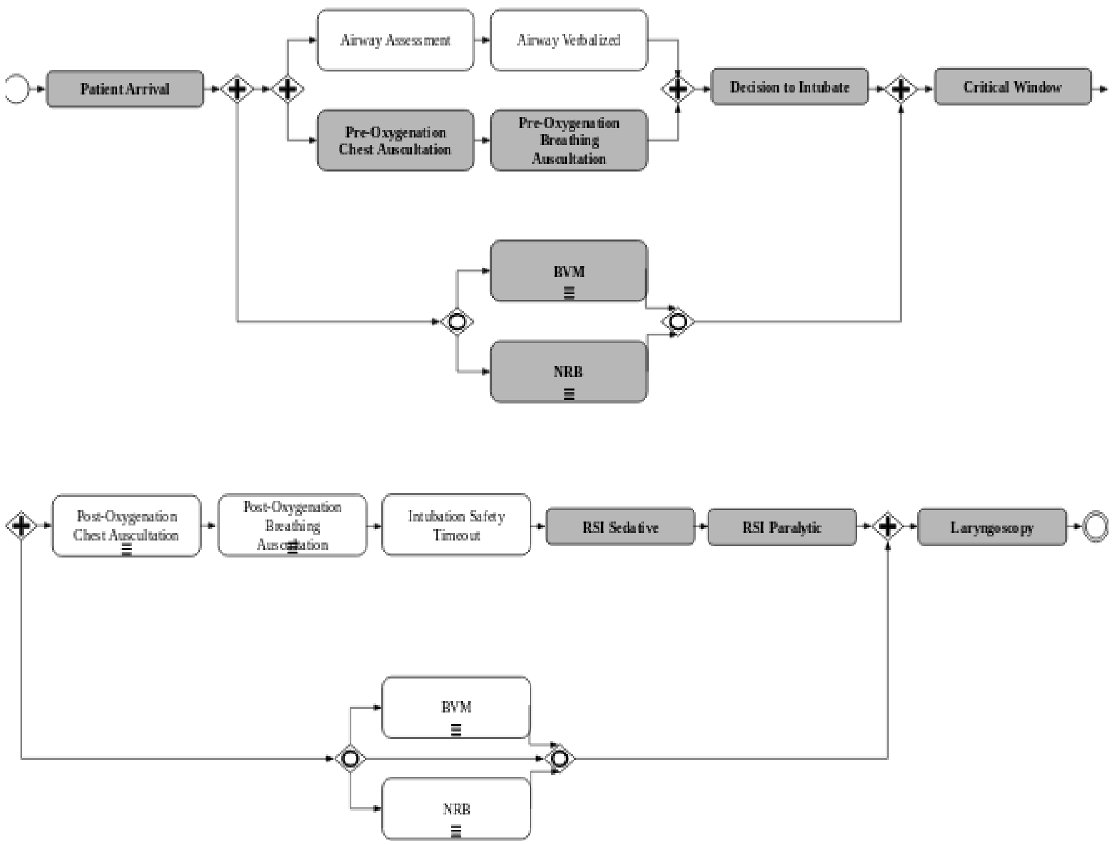


Figure 19 : BPMN diagram for Intubation highlighting the consensus activities for larger dataset

An illustration of the activities that appeared in the workflow model for span of 0 consecutive consensus activity, and the activities that didn't, is shown in figure 19. The highlighted activities in figure 19 corresponds to the activities that are consensus activities (backbone of the workflow model shown in figure 18). Figure 19 shows the visual contrast between the consensus activities (greyed out) and the non-consensus activities, but for a better quantitative analysis of the obtained results, we will use the Levenshtein distance, to determine the “distance” between the workflow model and the expert based model [7], like we did in the previous sections. Our assumption that the three operations (deletions, insertions and substitutions) that contribute to the Levenshtein distance are of equal cost (i.e., 1 unit cost operation for each deletions, insertions and substitutions) still holds for the following analysis, and the calculations are done exactly the same way as it was in the previous sections.

In figure 18, for span of 0 consecutive consensus activity, the workflow model needs insertion of 'Airway Assessment', 'Airway Verbalised', 'Post Oxygenation Chest Auscultation', 'Post Oxygenation Breathing Auscultation', 'Intubation Safety Timeout' and either of 'BVM' or 'NRB' (after 'Critical Window'). This means that are 6 insertions required, which makes the Levenshtein distance equal to 6-unit operations. Lesser the Levenshtein distance, means the "distance" between the workflow model and the expert-based model is lesser and thus more similarity between the workflow and the established "ground truth". We calculate the Levenshtein distance for each of the span values and determine the optimum value at which the Levenshtein distance is least and the workflow model exhibits most precise flow of treatment activities.

9.3. Analysis Using Larger Intubation Dataset for Span of 1 consecutive consensus activity

Continuing our analysis for Span of 1 consecutive consensus activity, we might get additional common but dispersed activities, if there are non-consensus activities spread across a span of 1 consensus activity, that has combined probability of occurrence more than the threshold (0.37). For span of 1 consecutive consensus activity we get the following diagram shown in figure 20 -

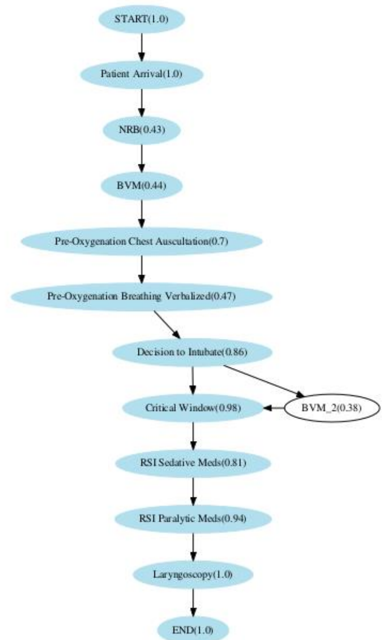


Figure 20 : Workflow model for larger dataset at span of 1 consecutive consensus activity

Figure 20 shows the workflow model for span of 1 consecutive consensus activity that has a branch, unlike the linear workflow model shown in figure 18 for the span of 0 consecutive consensus activity that was an oversimplification resulted from the short span (i.e., zero). From figure 20, we get 'BVM' as the common but dispersed activity. However, since we already have a BVM in backbone that corresponds to the BVM that appears before 'Critical Window', it appears to be a trivial addition to the workflow (appearing due to this particular alignment of BVM in the alignment matrix), but it couldn't be classified as noise, because if we compare the workflow model with the expert based mode, we see that the order of appearance is correct for BVM, as BVM can occur between 'Decision to Intubate' and 'Critical Window'. The workflow model needs insertion of 'Airway Assessment', 'Airway Verbalised', 'Post Oxygenation Chest Auscultation', 'Post Oxygenation Breathing Auscultation', 'Intubation Safety Timeout' and either of ‘BVM’ or ‘NRB’ (after ‘Critical Window’). This means that are 6 insertions required, which makes the Levenshtein distance equal to 6-unit operations. Thus, we can conclude that for span of 1 consecutive consensus activity we don't have an improvement over span of 0 consecutive consensus activity - the model stays similar i.e. deficient of crucial treatment activities, with a trivial addition.

9.4. Analysis Using Larger Intubation Dataset for Span of 2 consecutive consensus activities

Continuing our analysis for the span of 2 consecutive consensus activities, we get the following model, in figure 21 -

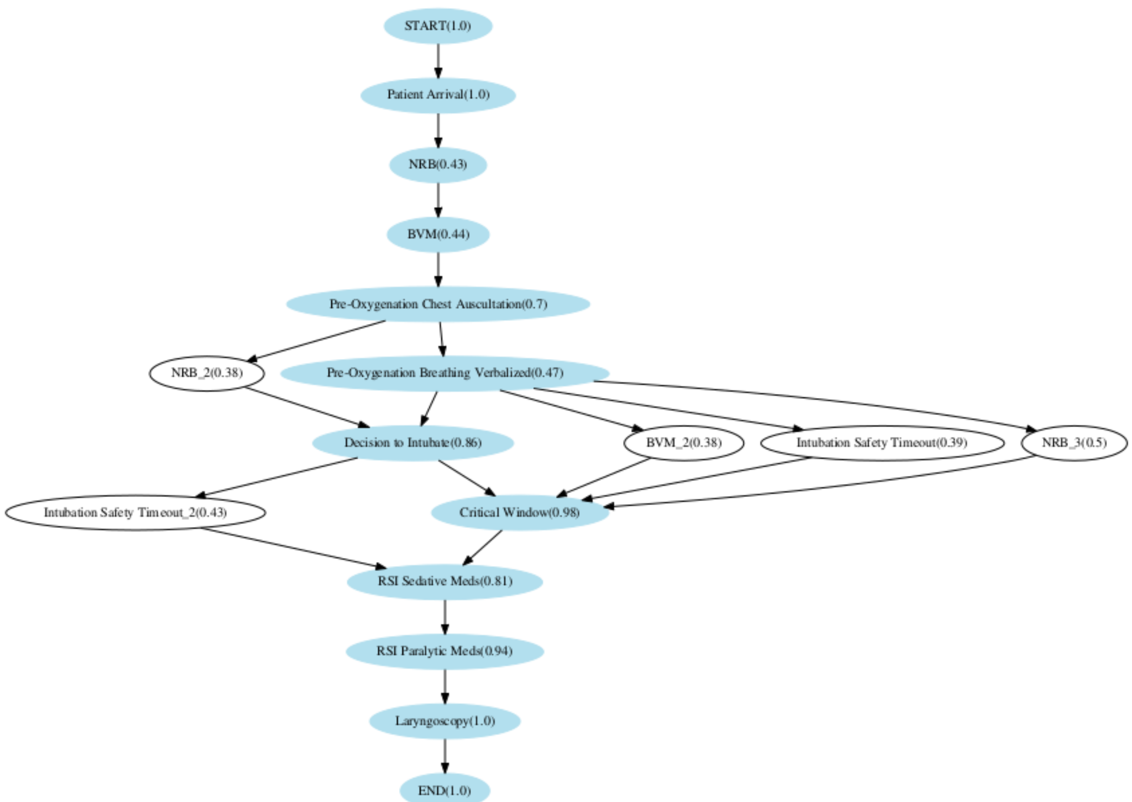


Figure 21 : Workflow model for larger dataset at span of 2 consecutive consensus activities

From figure 21, it is evident that the model has more branches than we had for span of 1 consecutive consensus activity. These branches corresponds to more common but dispersed activities introduced in the workflow for the span of 2 consecutive consensus activities, than we had for span of 1 consecutive consensus activity, but with these additional activities, we need to compare the workflow model with the expert model (figure 3) to check if these are relevant to the procedure or are just added noise. The model has significant additions of common but dispersed activities, as compared to the previous models. At this point, we have ‘Intubation Safety Timeout’ that appears between ‘Pre-Oxygenation Breathing Verbalised’ and ‘Critical WIndow’ and ‘Intubation Safety Timeout_2’ appearing between ‘Decision to Intubate’ and ‘RSI Sedative Meds’. On comparison with the expert-based model in figure 3, we see that ‘Intubation Safety Timeout’ neither appears before the ‘Critical Window’, nor the ‘RSI Sedative Meds’, as shown in figure 21. Thus, these two treatment activities of ‘Intubation Safety Timeout’ could be classified as noise. We also see the addition of BVM_2, NRB_2 and NRB_3 to the workflow. Given that we already have a BVM, and NRB in the workflow corresponding to the BVM and NRB that appear before the Critical Window - these additions (BVM_2, NRB_2 and NRB_3) could be classified as trivial additions to workflow, that aren’t noise since they appear in the correct order before Critical Window. Summarizing our analysis - we get 2 noises (‘Intubation Safety Timeout’ and ‘Intubation Safety Timeout_2’) and 3 trivial (BVM_2, NRB_2 and NRB_3) additions of treatment activities. For the calculation of Levenshtein distance, we need the deletion of ‘Intubation Safety Timeout’ and ‘Intubation Safety Timeout_2’ that appears as noise. The workflow model still needs insertion of ‘Airway Assessment’, ‘Airway Verbalised’, ‘Post Oxygenation Chest Auscultation’, ‘Post Oxygenation Breathing Auscultation’, ‘Intubation Safety Timeout’ and either of ‘BVM’ or ‘NRB’ (after ‘Critical Window’). This means that are 6 insertions and 2 deletions required, which makes the Levenshtein distance equal to 8-unit operations.

9.5. Analysis Using Larger Intubation Dataset for Span of 3 consecutive consensus activities

The model generated for span of 3 consecutive consensus activities, is shown below in figure 22 -

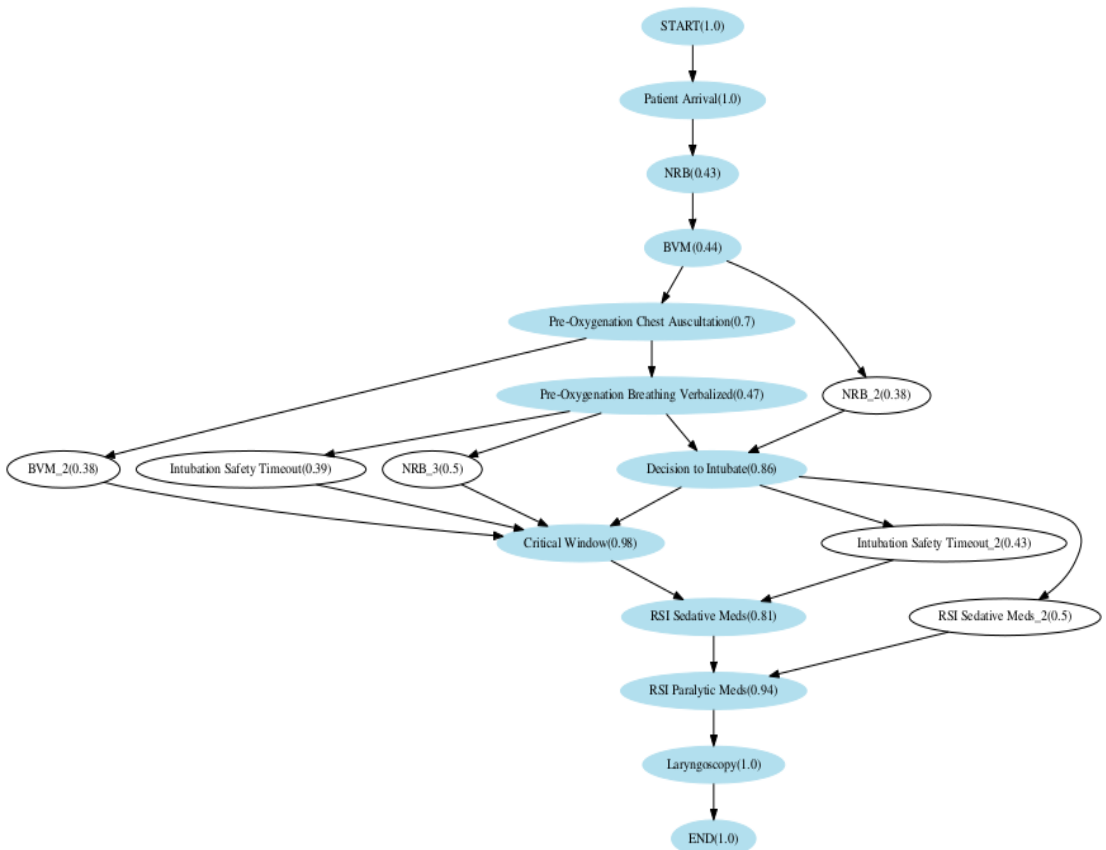


Figure 22 : Workflow model for larger dataset at span of 3 consecutive consensus activities

From figure 22, it is clear that we have additional branches, that corresponds to more common but dispersed activities as compared to the model generated for span of 2 consecutive consensus activities. The model shown in figure 22 is different from the model generated for span of 2 consecutive consensus activities, in NRB_2 and BVM_2 being reordered, and an additional ‘RSI Sedative Meds_2’ in the workflow. We still have ‘Intubation Safety Timeout’ that appears between ‘Pre-Oxygenation Breathing Verbalised’ and ‘Critical WIndow’ and ‘Intubation Safety Timeout_2’ appearing between ‘Decision to Intubate’ and ‘RSI Sedative Meds’. On comparison with the expert-based model in figure 3, we see that ‘Intubation Safety Timeout’ neither appears before the ‘Critical Window’, nor the ‘RSI Sedative Meds’, as shown in figure 22. Thus, these two treatment activities of ‘Intubation Safety Timeout’ could be classified as noise. Since, we already have established that BVM, and NRB are the trivial additions to the workflow in our discussion for the span of 2 consecutive consensus activities, (owing it to their already existing representation in backbone of the workflow) the reordering of NRB_2 and BVM_2 does not change the fact that they still hold the trivial additions status, as they aren't noise since they appear in the correct order before Critical Window. Again for this case, BVM_2, NRB_2, and NRB_3 and could be seen as the trivial additions to workflow, that aren’t noise, but ‘RSI Sedative Meds_2’ could be classified as noise because of its placement between ‘Decision to Intubate’ and ‘RSI Paralytic Meds’. Given that we already have the correct representation of ‘RSI Sedative Meds’ in our backbone, the ‘RSI Sedative Meds_2’ appear right after ‘Decision to Intubate’ which is not a correct representation of the flow when compared with ‘RSI Sedative Meds’ that appears more in-order (between ‘Critical Window’ and ‘RSI_Sedative Meds’ as should have). Summarizing our analysis - we get 3 noises (‘Intubation Safety Timeout’, ‘Intubation Safety Timeout_2’ and ‘RSI Sedative Meds_2’) and 3 trivial additions (BVM_2, NRB_2, and NRB_3) of treatment activities. Theoretically, the model generated for span of 3 consecutive consensus activities, has not been an improvement over span of 2 consecutive consensus activities, due to an increased noise. For the calculation of Levenshtein distance, we need the deletion of the 3 noises i.e., 'Intubation Safety Timeout', 'Intubation Safety Timeout_2' and 'RSI Sedative Meds_2'. The workflow model still needs insertion of 'Airway Assessment', 'Airway Verbalised', 'Post Oxygenation Chest Auscultation', 'Post Oxygenation Breathing Auscultation', 'Intubation Safety Timeout' and either of 'BVM' or 'NRB' (after 'Critical Window'). This means that are 6 insertions and 3 deletions required, which makes the Levenshtein distance equal to 9-unit operations.

9.6. Analysis Using Larger Intubation Dataset for Span of 4 consecutive consensus activities

The model generated for span of 4 consecutive consensus activities is shown below in figure 23 -

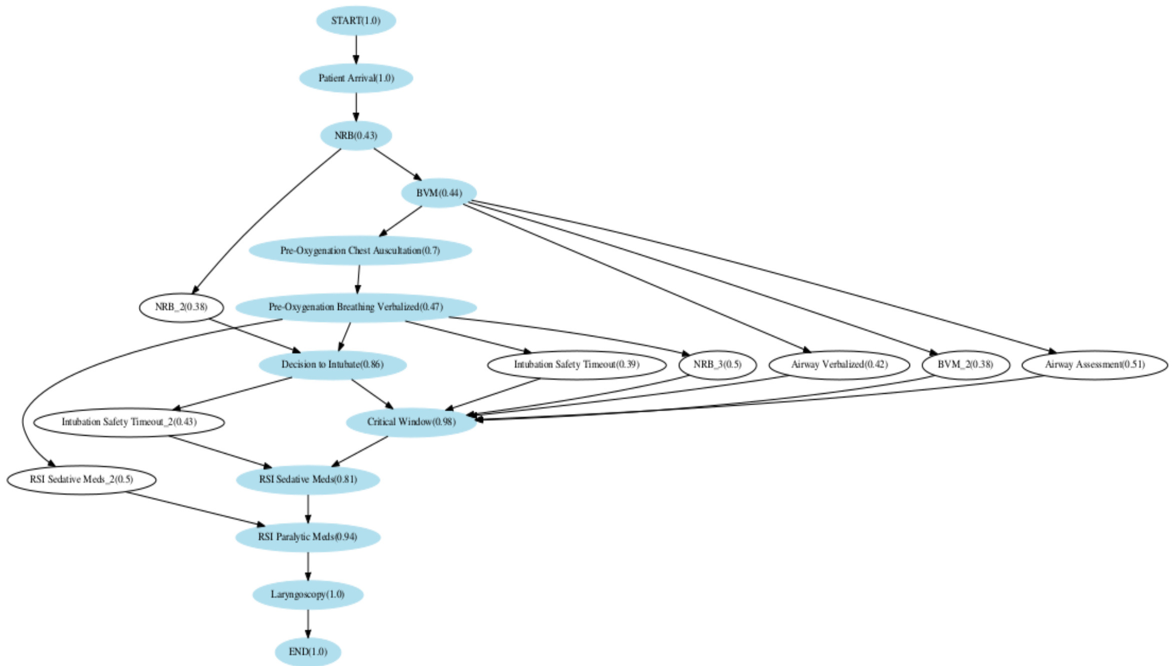


Figure 23 : Workflow model for larger dataset at span of 4 consecutive consensus activities

The model shown in figure 23, introduces ‘Airway Assessment’ and ‘Airway Verbalized’ before the ‘Critical Window’ to the workflow, that increases the accuracy of the model since it adds the two required insertions of ‘Airway Assessment’ and ‘Airway Assessment’, and in correct order (i.e. between ‘BVM’ and ‘Critical Window’. The model also shows that NRB_2 is reordered, but it means the same as before, since it was a trivial addition our workflow. RSI Sedative Meds_2 appears between ‘Pre-Oxygenation Breathing Verbalised and ‘RSI Paralytic Meds’, which is out of order, and is classified as noise (as discussed in previous sections). BVM_2, NRB_2, and NRB_3 still are trivial additions to the workflow. Summarizing our analysis - we get 3 noises (‘Intubation Safety Timeout’, ‘Intubation Safety Timeout_2’ and ‘RSI Sedative Meds_2’), 3 trivial additions (BVM_2, NRB_2, and NRB_3) and 2 correct additions of treatment activities (‘Airway Assessment’ and ‘Airway Verbalized’). Theoretically, the model generated for span of 4 consecutive consensus activities has been an improvement over span of 3 consecutive consensus activities due to the correct addition of 2 common but dispersed activities. For the calculation of Levenshtein distance, we need the deletion of the 3 noises i.e., ‘Intubation Safety Timeout’, ‘Intubation Safety Timeout_2’ and ‘RSI Sedative Meds_2’. The workflow model still needs insertion of ‘Post Oxygenation Chest Auscultation’, ‘Post Oxygenation Breathing Auscultation’, ‘Intubation Safety Timeout’ and either of ‘BVM’ or ‘NRB’ (after ‘Critical Window’). This means that are 4 insertions and 3 deletions required, which makes the Levenshtein distance equal to 7-unit operations.

9.7. Analysis Using Larger Intubation Dataset for Span of 5 consecutive consensus activities

The workflow model for span of 5 consecutive consensus activities is shown below in figure 24 –

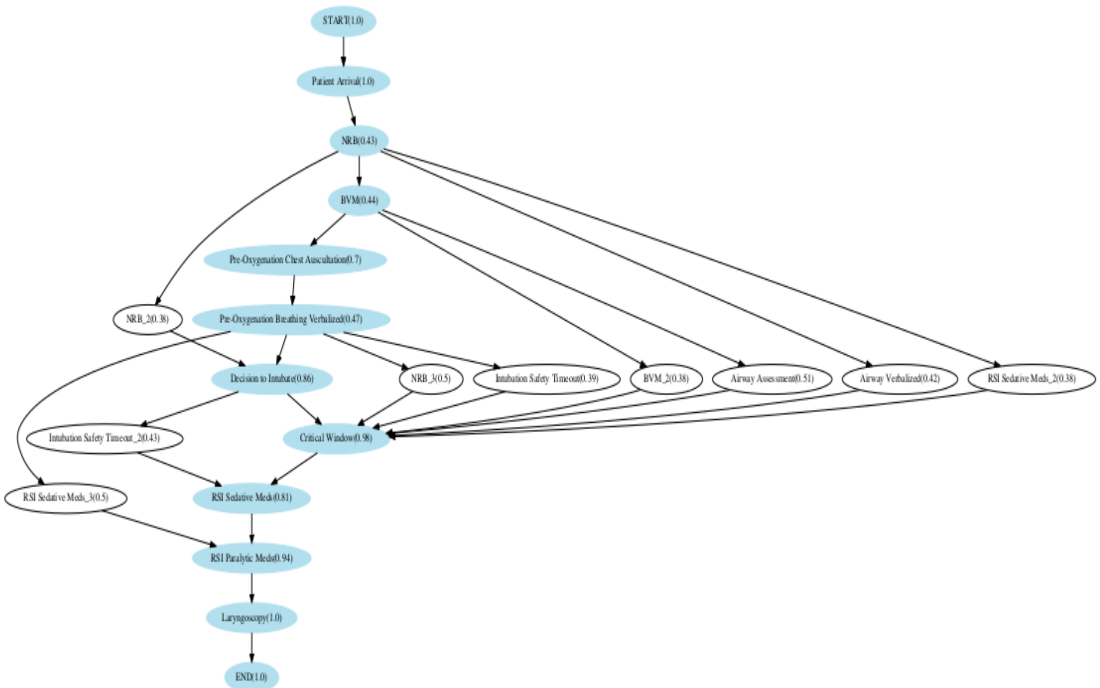


Figure 24 : Workflow model for larger dataset at span of 5 consecutive consensus activities

The model shown in figure 24, is similar to the model generated for span of 4 consecutive consensus activities, but with the addition of 'RSI Sedative Meds_3' that appears between 'Pre-Oxygenated Breathing Verbalized' and 'RSI Paralytic Meds'. Given that we already have the correct representation of 'RSI Sedative Meds' in our backbone, the 'RSI Sedative Meds_3' appears before 'RSI Paralytic Meds' which is a correct representation of the flow when compared with the expert model in figure 3. However since we already have a 'RSI Sedative Meds' that appears more in-order (between 'Critical Window' and 'RSI_Sedative Meds' as it should have - according to the expert model in figure 3), we can classify 'RSI Sedative Meds_3' as a trivial addition to the workflow. We also have 'RSI Sedative Meds_2' between the 'NRB' and 'Critical Window', and the reordered 'Airway Verbalised' between the same 'NRB' and 'Critical Window'. Comparing the workflow model in figure 24, we see that 'RSI Sedative Meds' is an activity that appears after ‘Critical Window’, however the ‘RSI Sedative Meds_2’ in the workflow appears between the ‘NRB’ and ‘Critical Window’, therefore indicating an incorrect flow. ‘RSI Sedative Meds_2’ could be thus classified as noise. Also, the order of occurrence of ‘Airway Verbalised’ was better depicted in the previous model (between ‘BVM’ and ‘Critical Window’). Due to the rearrangement in the flow of ‘Airway Verbalised’, according to figure 24 ‘Airway Verbalised’ could be carried out before ‘Airway Assessment’, which is not in agreement with the expert model in figure 3 (according to expert model, ‘Airway Verbalised’ occurs after ‘Airway Assessment’). Since the reordering of ‘Airway Verbalised’ has disturbed the flow of occurrence of both ‘Airway Assessment’ and ‘Airway Verbalised’, both of these activities have lost their preciseness in workflow order. Span of 5 consecutive consensus activities is not an improvement over span of 4 consecutive consensus activities, because of the disturbed rearrangement of Airway Verbalised and the added noise of 'RSI Sedative Meds_2’. For the calculation of Levenshtein distance, we need the deletion of the 3 noises i.e., ‘Intubation Safety Timeout’, ‘Intubation Safety Timeout_2’, and 'RSI Sedative Meds_2’. The workflow model still needs insertion of ‘Post Oxygenation Chest Auscultation’, ‘Post Oxygenation Breathing Auscultation’, ‘Intubation Safety Timeout’ and either of ‘BVM’ or ‘NRB’ (after ‘Critical Window’). Along with insertions and deletions, we address the lost accuracy in the order of occurrence ‘Airway Assessment’ and ‘Airway Verbalised’, (more apt representation in the previous model) and count them as the required substitutions. This means that there are 4 insertions, 3 deletions and 2 substitutions required, which makes the Levenshtein distance equal to 9-unit operations.

9.8. Analysis Using Larger Intubation Dataset for Span of 6 consecutive consensus activities

Continuing our analysis for span of 6 consecutive consensus activities, we get the following model in figure 25 -

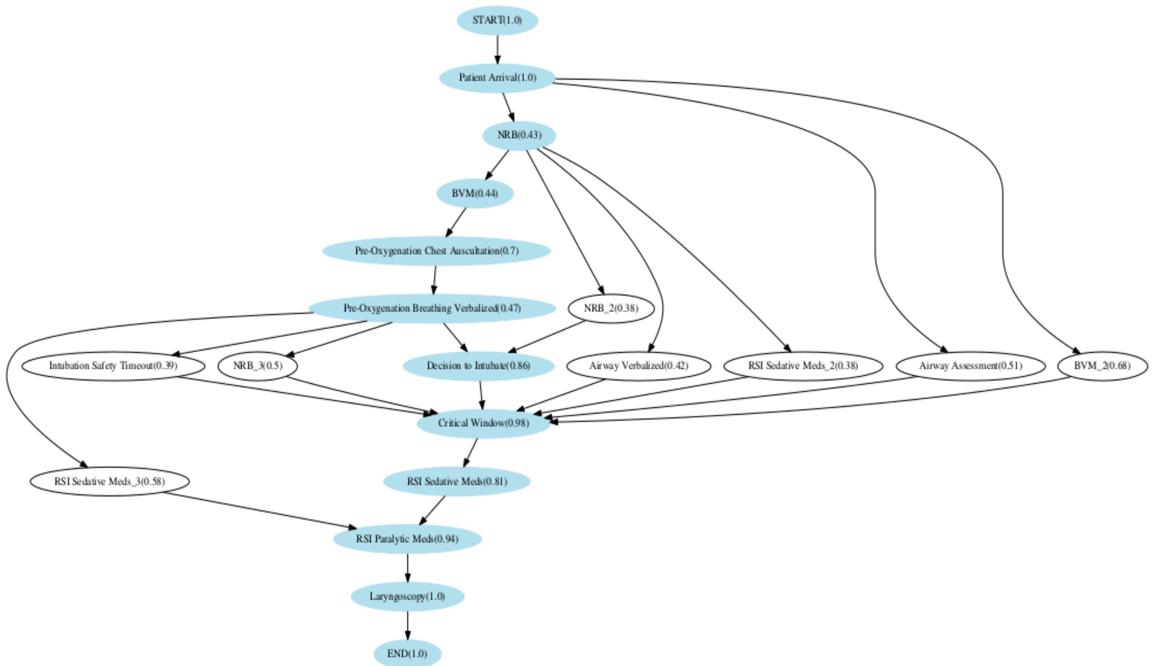


Figure 25 : Workflow model for larger dataset at span value 6

The model generated for span of 6 consecutive consensus activities fine tunes the occurrence of 'Airway Assessment' that now appears between 'Patient Arrival' and 'Critical Window'. The reordering of 'Airway Assessment' fine tunes the flow of activities, as it reorders 'Airway Assessment' to appear before 'Airway Verbalised' (as shown in the expert model in figure 3). Figure 25 also gets rid of 'Intubation Safety Timeout_2', that was classified as noise in the previous models. The model generated for span of 6 consecutive consensus activities could be seen as an improvement over the previous models, with 2 correct additions of common but dispersed activities ('Airway Assessment' and 'Airway Verbalised'), 2 noises ('Intubation Safety Timeout' and 'RSI Sedative Meds_2') and 4 trivial activities ('RSI Sedative Meds_3', BVM_2, NRB_2, and NRB_3). For the calculation of Levenshtein distance, we need the deletion of the 2 noises i.e., ‘Intubation Safety Timeout’, and ‘RSI Sedative Meds_2’. The workflow model still needs insertion of ‘Post Oxygenation Chest Auscultation’, ‘Post Oxygenation Breathing Auscultation’, ‘Intubation Safety Timeout’ and either of ‘BVM’ or ‘NRB’ (after ‘Critical Window’). This means that there are 4 insertions and 2 deletions, which makes the Levenshtein distance equal to 6-unit operations.

9.9. Analysis Using Larger Intubation Dataset for Span of 7 consecutive consensus activities

The workflow model for span value of 7 consecutive consensus activities is shown below in figure 26 -

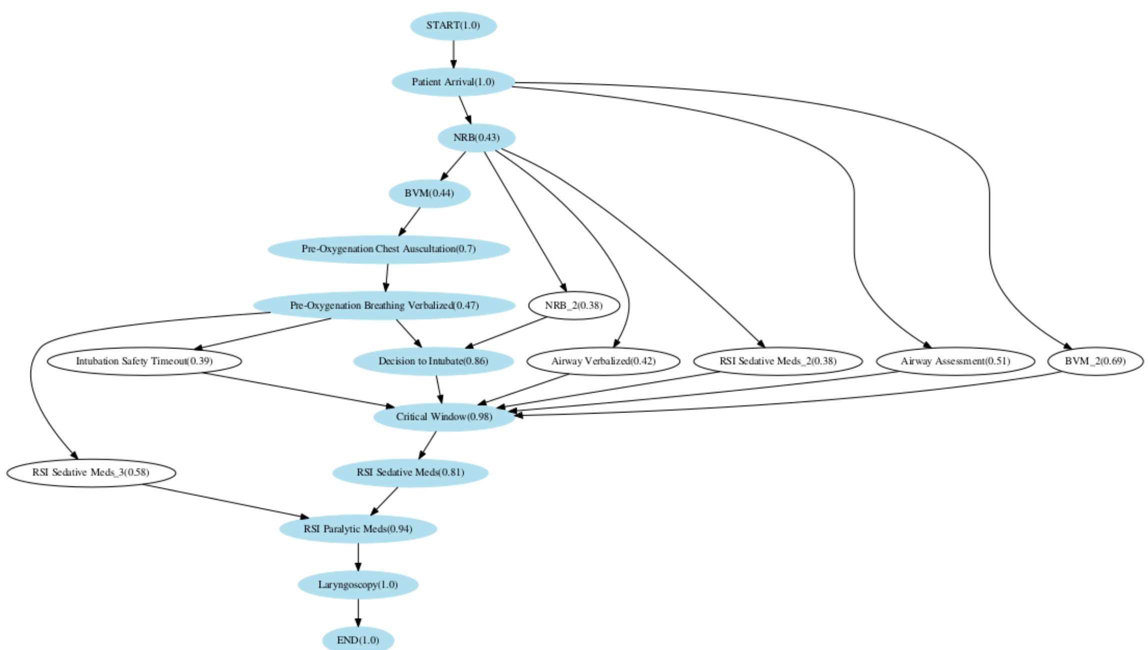


Figure 26 : Workflow model for larger dataset at span of 7 consecutive consensus activities

For the span value of 7 consecutive consensus activities, we get the same model, but without ‘NRB_3’, that was a trivial addition to the workflow (as discussed in previous sections). For the calculation of Levenshtein distance, we need the deletion of the 2 noises i.e., ‘Intubation Safety Timeout’, and ‘RSI Sedative Meds_2’. The workflow model still needs insertion of ‘Post Oxygenation Chest Auscultation’, ‘Post Oxygenation Breathing Auscultation’, ‘Intubation Safety Timeout’ and either of ‘BVM’ or ‘NRB’ (after ‘Critical Window’). This means that there are 4 insertions and 2 deletions, which makes the Levenshtein distance equal to 6-unit operations. In terms of reduced complexity, we would prefer the span value of 7 consecutive consensus activities, with lesser activities giving the same amount of information as the previous model.

9.10. Analysis Using Larger Intubation Dataset for Span of 8 consecutive consensus activities

The workflow model generated for the span value of 8 consecutive consensus activities is shown below in figure 27 -

Figure 27 : Workflow model for larger dataset at span of 8 consecutive consensus activities

The model generated for the span value of 8 consecutive consensus activities shown in figure 27, is the same as what we got for span of 7 consecutive consensus activities. Hence, our analysis remains the same as it was for span of 7 consecutive consensus activities, and the Levenshtein distance is equal to 6-unit operations.

9.11. Analysis Using Larger Intubation Dataset for Span of 9 consecutive consensus activities

Moving to the span value of 9 consecutive consensus activities, we get the following diagram in figure 28 -

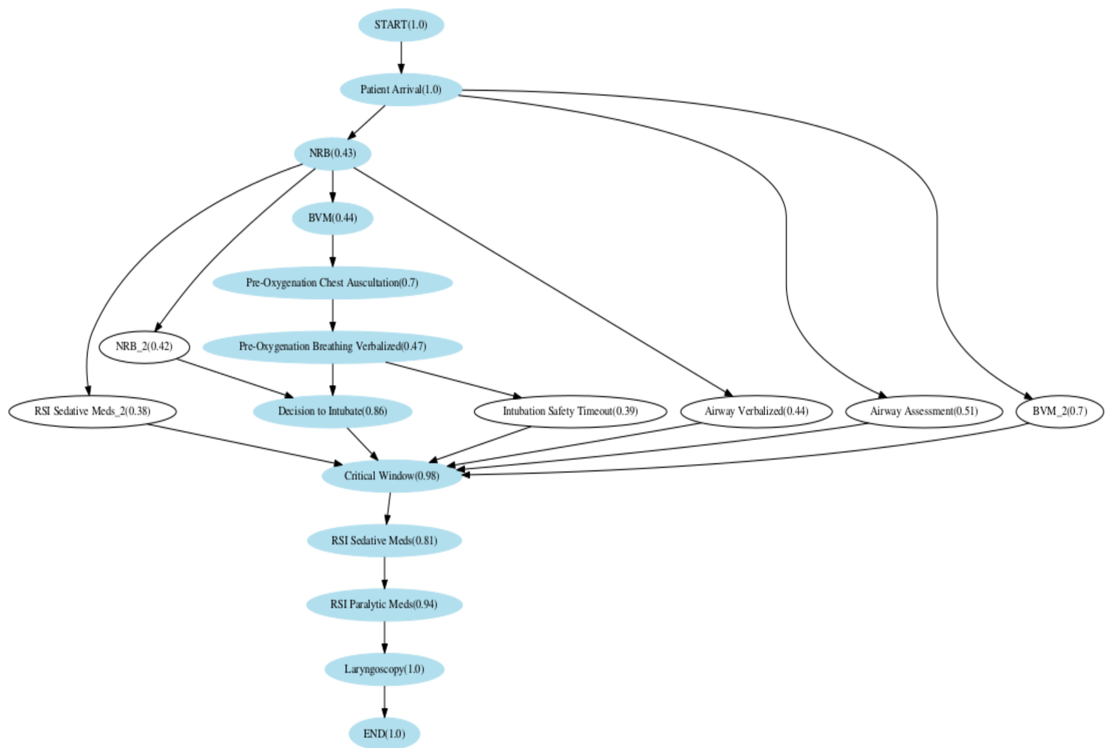


Figure 28 : Workflow model for larger dataset at span of 9 consecutive consensus activities

The model shown in figure 28 is rather similar to the previous model, but with the riddance of 'RSI Sedative Meds_3' that was a trivial addition to the workflow. We still have 'RSI sedative Meds_2' and 'Intubation Safety Timeout' in the workflow, as noises from the previous models. For the calculation of Levenshtein distance, we need the deletion of the 2 noises i.e., 'Intubation Safety Timeout', and 'RSI Sedative Meds_2'. The workflow model still needs insertion of 'Post Oxygenation Chest Auscultation', 'Post Oxygenation Breathing Auscultation', 'Intubation Safety Timeout' and either of 'BVM' or 'NRB' (after 'Critical Window'). This means that there are 4 insertions and 2 deletions, which makes the Levenshtein distance equal to 6-unit operations. Span of 7 consecutive consensus activities, had an additional trivial activity ('RSI Sedative Meds_3) as compared to Span of 9 consecutive consensus activities, thus making span of 9 consecutive consensus activities more apt in terms of reduced complexity (we would prefer the span value of 9 consecutive consensus activities, with lesser activities giving the same amount of information as the previous model thereby reducing the model complexity).

9.12. Analysis Using Larger Intubation Dataset for Span of 10 consecutive consensus activities

The workflow models generated for span value of 10 consecutive consensus activities is shown below in figure 29 -

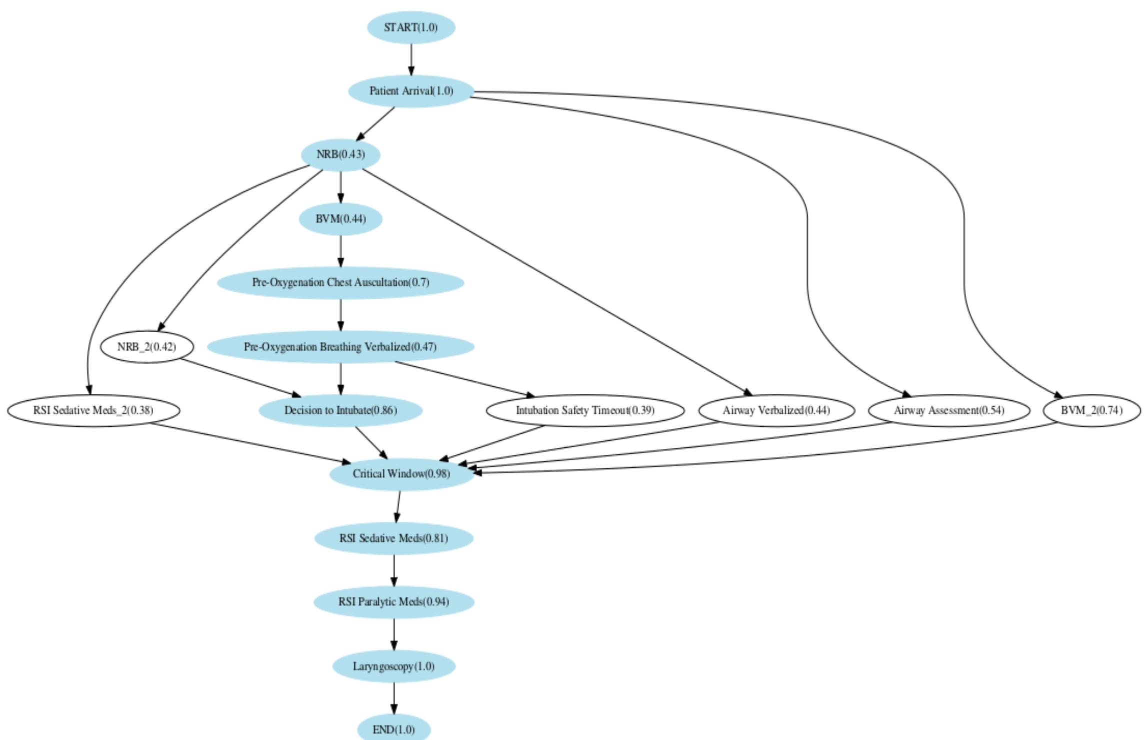


Figure 29 : Workflow model for larger dataset at span of 10 consecutive consensus activities

The models generated span value of 10 consecutive consensus activities is exactly the same as the model generated for span value of 9 consecutive consensus activities, and the Levenshtein distance is still equal to 6 unit operations (as for the span of 9 consecutive consensus activities) hence our analysis remains the same for both, and span of 9 consecutive consensus activities, still hold the place of the optimum value of span for this case.

Given that we have 10 consensus activities (excluding start and end), forming the backbone of our workflow model, our span analysis ranges only from 0 to 10 consecutive consensus activities. Thus, if we go for span values beyond 10 consecutive consensus activities, the model remains the same. For the span value of 11 consecutive consensus activities, we get the model shown below in figure 30 that is exactly the same model as we got for span of 10 consecutive consensus activities, hence the analysis remains the same -

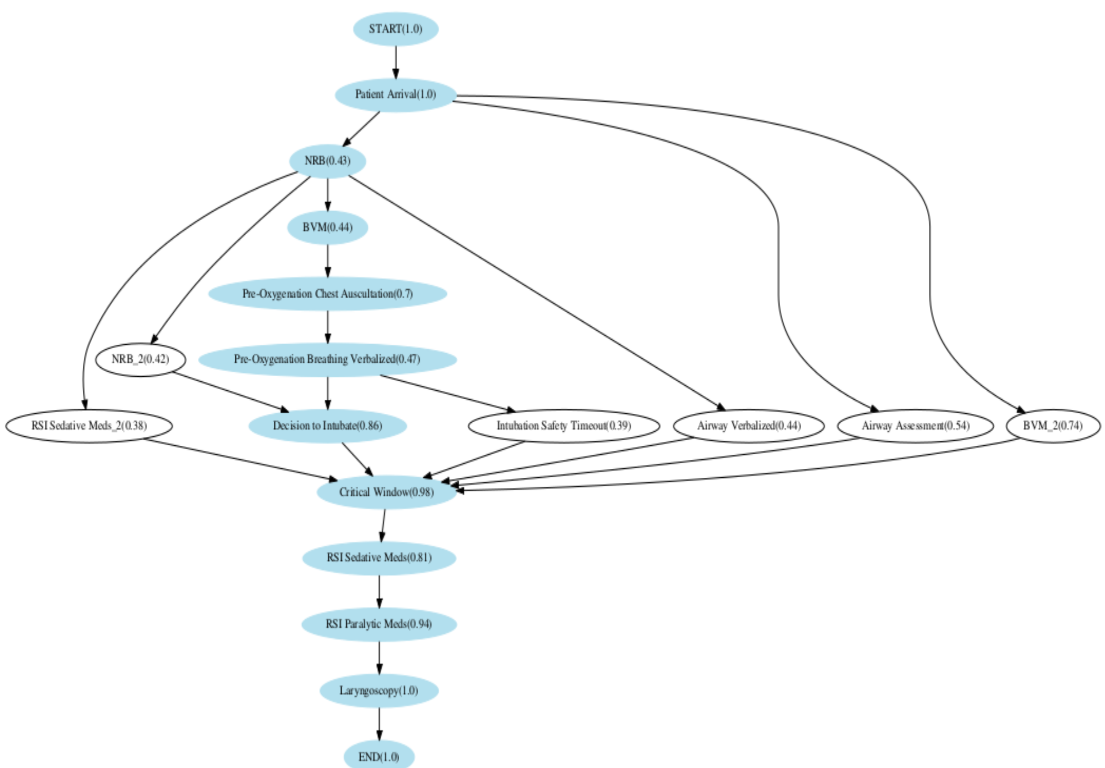


Figure 30 : Workflow model for larger dataset at span of 11 consecutive consensus activities

10. Results for Larger Intubation Dataset

Thus, for the given dataset, we have seen the varying accuracy and noise in the generated workflow model, at different values of span, ranging from span values of 0 to 10 consecutive consensus activities. We can conclude that, for this particular case, Span of 9 consecutive consensus activities turns out to be the optimum value of span, because for span of 9 consecutive consensus activities we get a model closest to the expert model, with the least Levenshtein distance, least amount of noise and least complexity (trivial additions).

For the models generated at span of 6 - 10 we got the same Levenshtein distance, but we chose the span of 9 consecutive consensus activities as an optimum value, because we get the least number of trivial additions in the workflow model for span of 9 consecutive consensus activities. We get the same results for span of 9 and 10 consecutive consensus activities, yet we choose span of 9 consecutive consensus activities as the optimum value because of the time complexity consideration. For the span value of 10 consecutive consensus activities, we get an iteration more through the alignment matrix as compared to the span value of 9 consecutive consensus activities, thus increasing the run time on a large dataset. Though the difference is less here, if we had a larger alignment matrix, the difference would have been significant. Hence, in such cases, it would be preferable to consider the run time complexity of the algorithm at the span values in consideration.

A more clear picture of the discussion could be seen in the condensation of our analysis for this particular case of Intubation procedure, shown in the table 10.1 that compares different values of span, to the changes in accuracy (inferred by Correct activities and Noise column) and complexity (inferred by Trivial activities and Noise) -

Span (of consecutive consensus activities)	**Number of consensus activities**	**Total number of activities**	**Correct activities (common but dispersed)**	**Trivial activities**	**Noise**	**Levenshtein Distance**
0	10	10	0	0	0	6
1	10	11	0	1	0	6
2	10	15	0	2	3	8
3	10	16	0	3	3	9
4	10	18	2	3	3	7
5	10	19	0	3	6	9
6	10	18	2	4	2	6
7	10	17	2	3	2	6
8	10	17	2	3	2	6
9	**10**	**16**	**2**	**2**	**2**	**6**
10	10	16	2	2	2	6

Table 10.1 : Workflow composition at different values of span (larger intubation dataset)

11. Conclusion

Throughout our analysis, we see that the varying values of span greatly affects the complexity of a workflow model. We get different optimum values of span for different dataset, because the alignment of activities depends upon the different cases that contributes to a dataset. Although we cannot generalize the optimum value of span, we have seen in Table 7.1 and Table 10.1, that the initial hypothesis that changing the hyperparameter Span, results in varying workflow configurations and it plays a crucial role in determining the accuracy of a workflow. For an even larger dataset, of a more complex medical procedure than intubation, we can get an easy interpretable workflow, by using span - ensuring the accuracy of procedural activities. The optimum value of Span is important to be studied and narrowed down, to vouch for the Workflow algorithm, as an improvement over the other workflow discovery algorithms.

12. References

1. National Practitioner Data Bank (2019): Adverse Action and Medical Malpractice Reports (1990 - March 31, 2019)
2. Makary Martin A, Daniel Michael. Medical error—the third leading cause of death in the US BMJ 2016; 353 :i2139
3. Demetriades D, Kimbrell B, Salim A, Velmahos G, Rhee P, Preston C, Gruzinski G, Chan L. Trauma deaths in a mature urban trauma system: is "trimodal" distribution a valid concept? J Am Coll Surg. 2005 Sep;201(3):343-8. PubMed PMID: 16125066.
4. Gruen RL, Jurkovich GJ, McIntyre LK, Foy HM, Maier RV. Patterns of errors contributing to trauma mortality: lessons learned from 2,594 deaths. Ann Surg. 2006 Sep;244(3):371-80. PubMed PMID: 16926563; PubMed Central PMCID: PMC1856538
5. Castaneda C, Nalley K, Mannion C, et al. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. *J Clin Bioinforma*. 2015;5:4. Published 2015 Mar 26. doi:10.1186/s13336-015-0019-3
6. Li, Jingyuan & Yang, Sen & Chen, Shuhong & Tao, Fei & Marsic, Ivan & Burd, Randall. (2018). Discovering Interpretable Medical Workflow Models. IEEE International Conference on Healthcare Informatics. IEEE International Conference on Healthcare Informatics. 2018. 437-439. 10.1109/ICHI.2018.00089.
7. Bouarfa L, Dankelman J. Workflow mining and outlier detection from clinical activity logs. J Biomed Inform. 2012 Dec;45(6):1185-90. PubMed PMID: 22925724