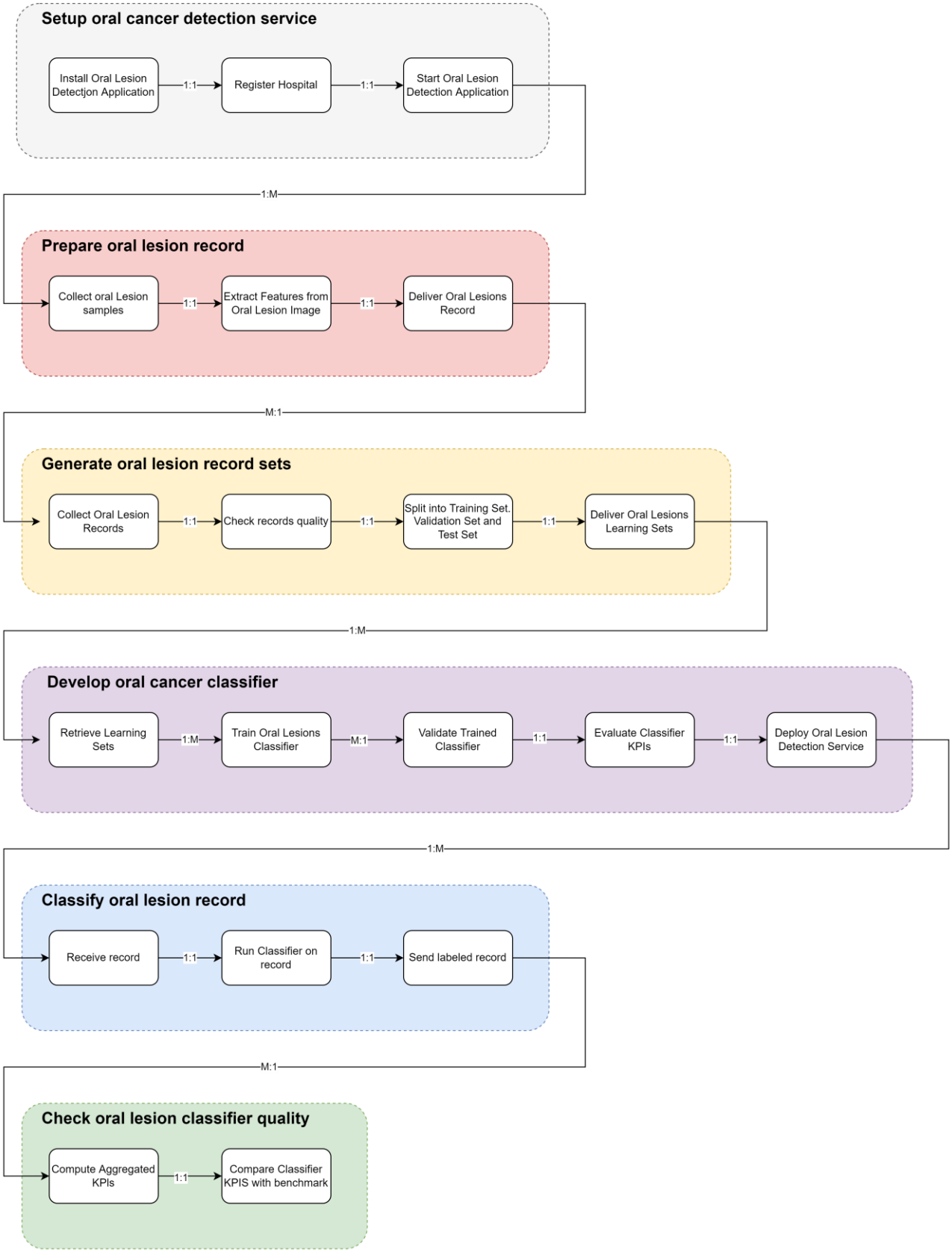
University of Pisa
Department of Information Engineering
Process Mining and Intelligence

Oral Lesion Detection Service

Arancio Febbo Salvatore, Di Donato Mattia, Martino Giuseppe, Giorgi Matteo

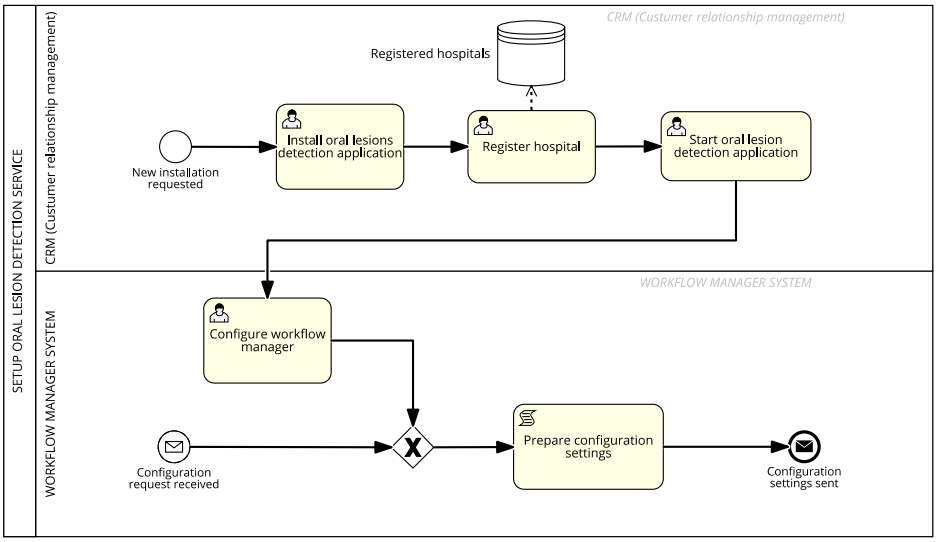
Contents

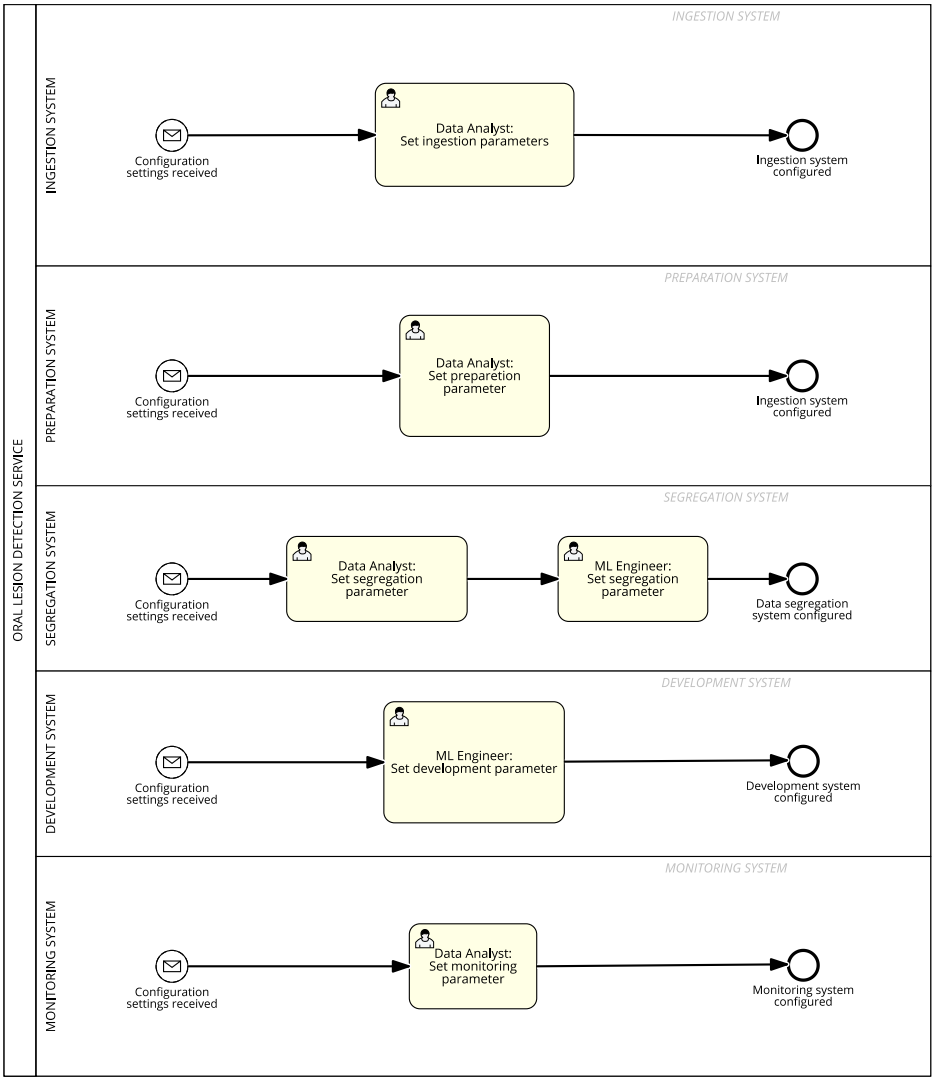
PROCESS LANDSCAPE



HANDOFF>SERVICE MODEL

Oral lesion detection configuration





New installation requested: receives the request for a new hospital that wants to be added to the system

Configuration request received: receives messages from individual tasks and the workflow manager system takes care of routing the request to the correct system.

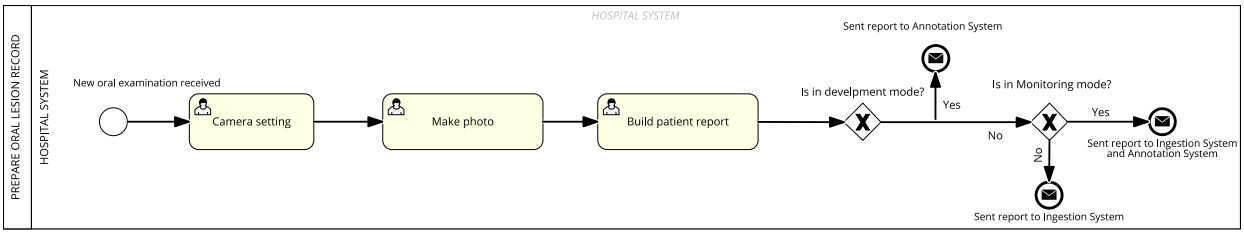
Configuration setting sent: takes care of forwarding the message to the correct system

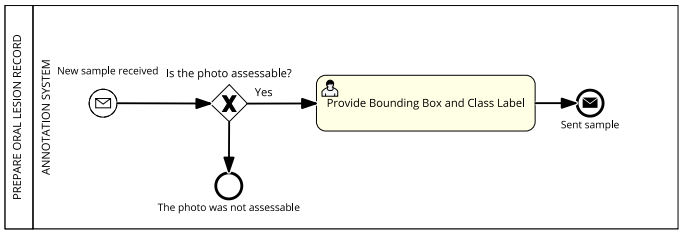
Configuration setting received: receives the message from the workflow manager system

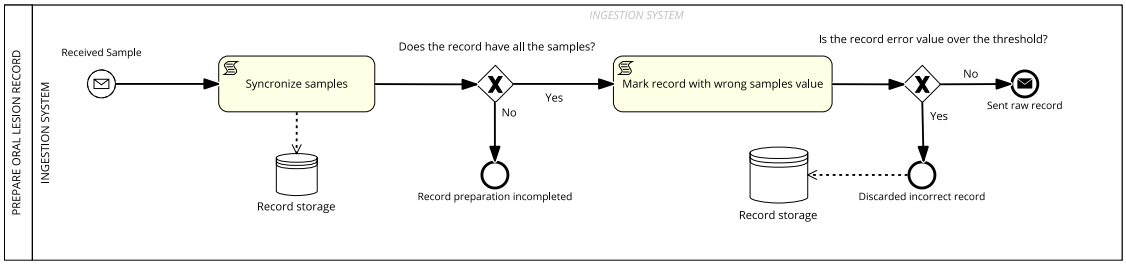
All the end events of the systems take care of forwarding messages to the interested processes.

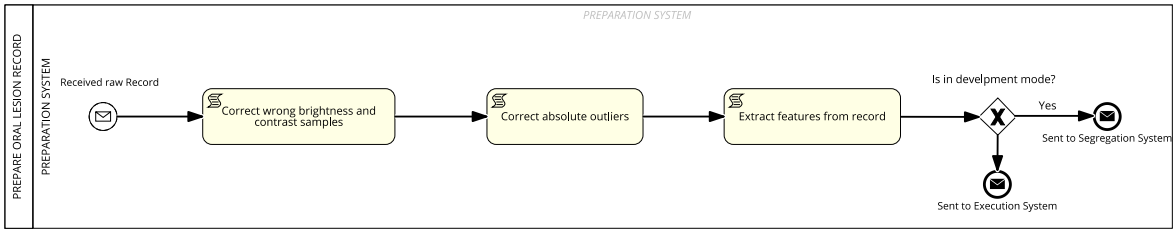
Prepare Oral Lesion Record (Mattia)

These systems exchange samples to create a record, which is the fundamental building block of our storage. A record is composed of various samples, including photographs taken by a doctor, patient reports containing historical information, and class labels assigned by an annotator.

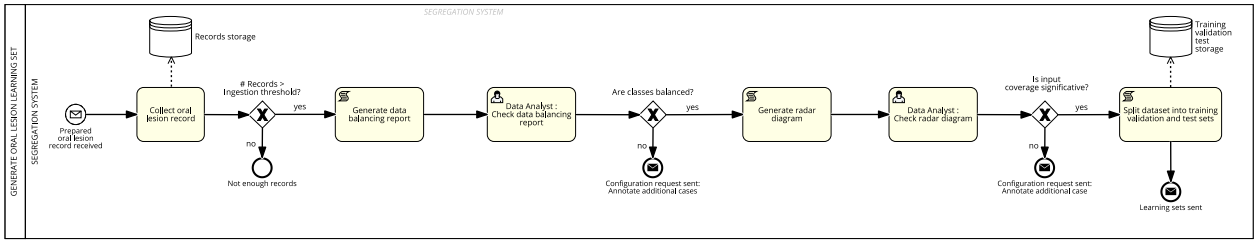




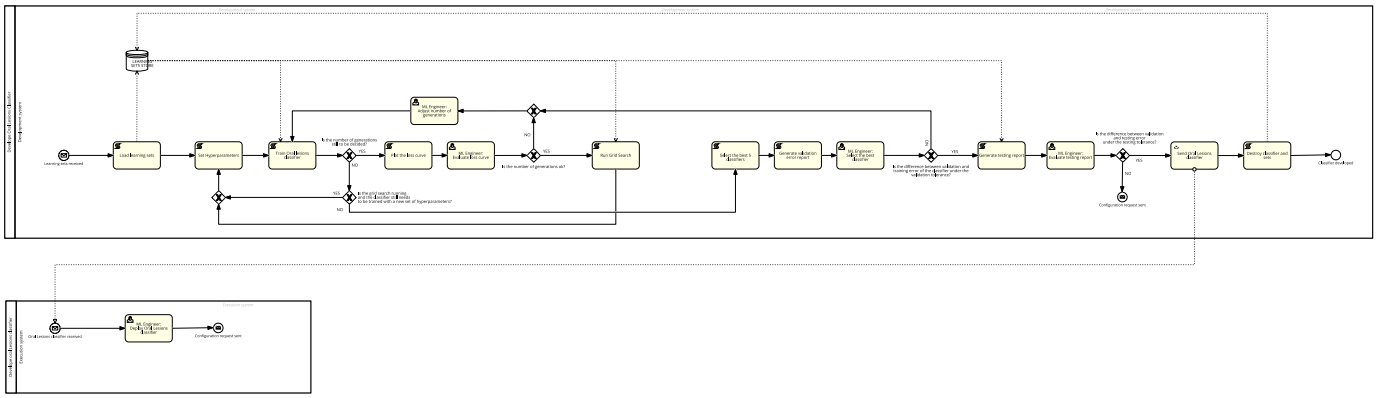




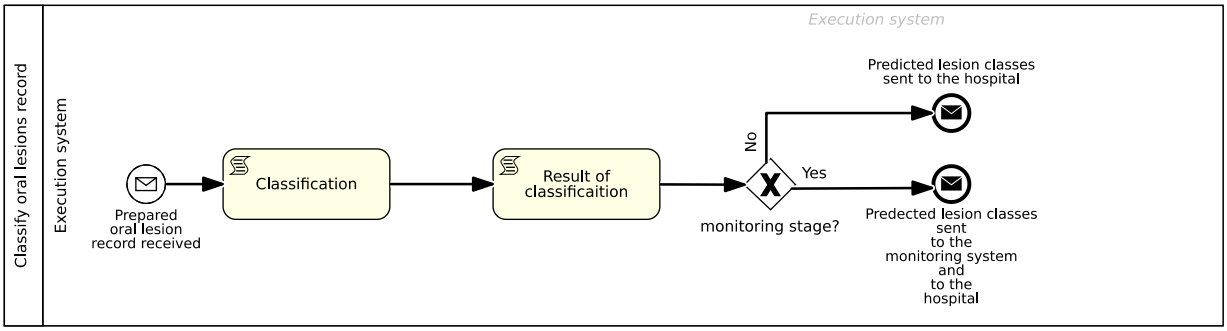
Generate Oral Lesion Learning Set (Matteo)



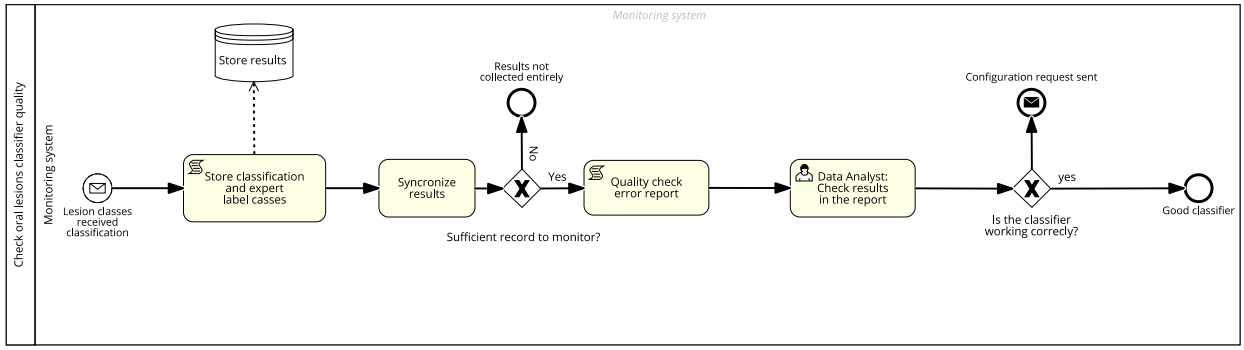
Develop Oral Lesion Classifier (Giuseppe)



Classify Oral Lesion record(Salvatore)



Check Oral Lesion Classifier Quality(Salvatore)



TASK LEVEL MODEL

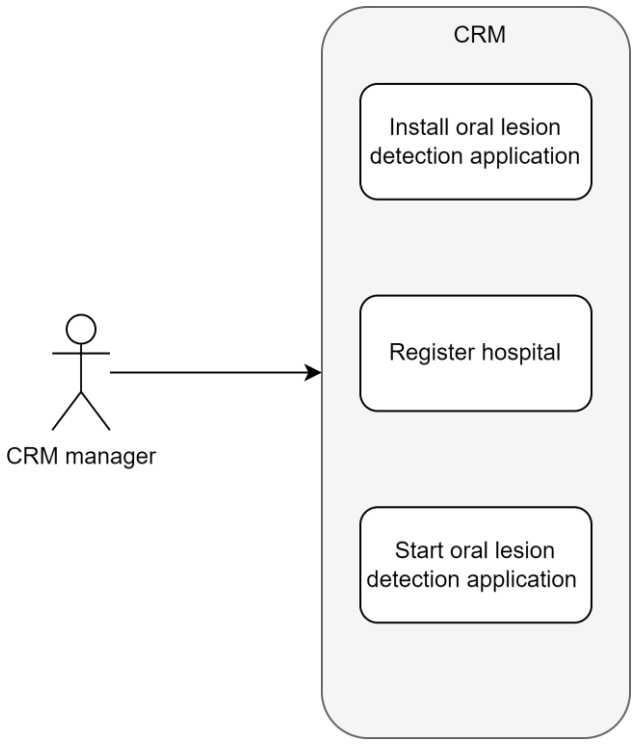
Human Actors' Salary (Matteo)

In the following table we estimate the salaries of the human actors of our factory.

Human Actor	Annual Salary	Unit Cost	Source
Customer relationship manager	£40000	40000/37000 =1,08	https://www.glassdoor.com/Salaries/uk-customer-relationship-manager-salary-SRCH_IL.0,2_IN2_KO3,32.htm?clickSource=searchBtn
Data analyst	£37000	37000/37000 =1,00	https://www.glassdoor.com/Salaries/uk-data-analyst-salary-SRCH_IL.0,2_IN2_KO3,15.htm?clickSource=searchBtn
ML engineer	£64000	64000/37000 =1,73	https://www.glassdoor.com/Salaries/uk-machine-learning-engineer-salary-SRCH_IL.0,2_IN2_KO3,28.htm?clickSource=searchBtn
Otolaryngologist	£77000	77000/37000 =2,08	https://www.glassdoor.com/Salaries/uk-medical-specialist-salary-SRCH_IL.0,2_IN2_KO3,21.htm?clickSource=searchBtn

USE CASE

CRM (Mattia Di Donato)



The CRM manager is the only human actor in the CRM system. His tasks are to install the application on the client's device, proceed with registration of the new hospital in the system with the insertion of the parameters. Finally launch the application. We computed the cost for this actor by dividing the average annual salary in UK by £40000.

	Year	Cost
CRM manager	£40000	40000 / 37000 = 1,08

Install oral lesion detection application (Mattia Di Donato)

In this phase we have these subtasks:

1. CRM Manager proceeds to install the application on the hospital machine.
2. **HOSPITAL MACHINE**: notifies the installation result.

Subtask	Actor	Cognitive Effort	Probability	Total Cost
Install the application	CRM Manager	1-Remember	100%	100%*1*1,08=1,08
Total Cost				1,08

Register hospital (Mattia Di Donato)

In this phase we have these subtasks:

1. CRM Manager open the application.
2. CRM Manager go to sign-up window.
3. CRM Manager insert the mean number of patients for year.
4. CRM Manager insert the mean percentage number of foreign patients.

5. CRM Manager insert the percentage number of generic lesion.
6. CRM Manager insert the percentage number of aphthous lesion.
7. CRM Manager insert the percentage number of neoplastic lesion.
8. CRM Manager insert the hospital dimension.
9. **APPLICATION**: notifies the registration result with the ID associated with the new hospital.
10. CRM Manager print the new network parameter.

Subtask	Actor	Cognitive Effort	Probability	Total Cost
Open the application	CRM Manager	1-Remember	100%	100%*1*1,08=1,08
Open sign-up application window	CRM Manager	1-Remember	100%	100%*1*1,08=1,08
Insert mean number of patients for year	CRM Manager	1-Remember	100%	100%*1*1,08=1,08
Insert the mean percentage number of foreign patients	CRM Manager	1-Remember	100%	100%*1*1,08=1,08
Insert the percentage number of generic lesion	CRM Manager	1-Remember	100%	100%*1*1,08=1,08
Insert the percentage number of aphthous lesion	CRM Manager	1-Remember	100%	100%*1*1,08=1,08
Insert the percentage number of neoplastic lesion	CRM Manager	1-Remember	100%	100%*1*1,08=1,08
Insert hospital dimension	CRM Manager	2-Understand	100%	100%*2*1,08=2,16
Print the network params	CRM Manager	1-Remember	100%	100%*1*1,08=1,08
Total Cost				11,88

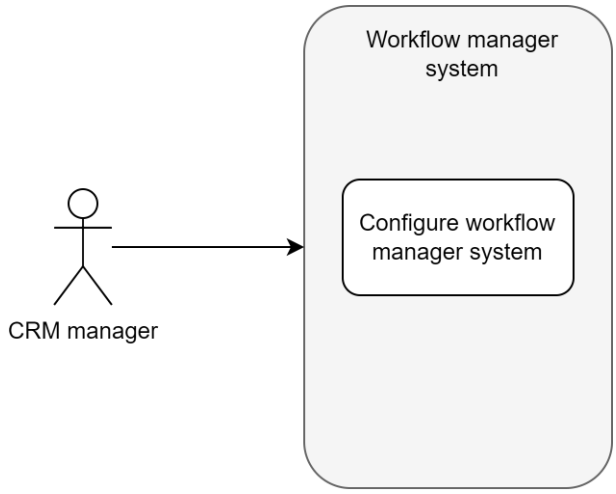
Start oral lesion detection application (Mattia Di Donato)

In this phase we have these subtasks:

1. CRM Manager starts the application in operative mode.
2. **APPLICATION**: notifies the operative mode activation.

Subtask	Actor	Cognitive Effort	Probability	Total Cost
Starts the application	CRM Manager	1-Remember	100%	100%*1*1,08=1,08
Total Cost				1,08

Workflow Manager System (Mattia Di Donato)



The CRM manager is the only human actor in the Workflow Manager System. His task is to configure the workflow manager system. We computed the cost for this actor by dividing the average annual salary in UK by £40000.

	Year	Cost
CRM manager	£40000	40000 / 37000 = 1,08

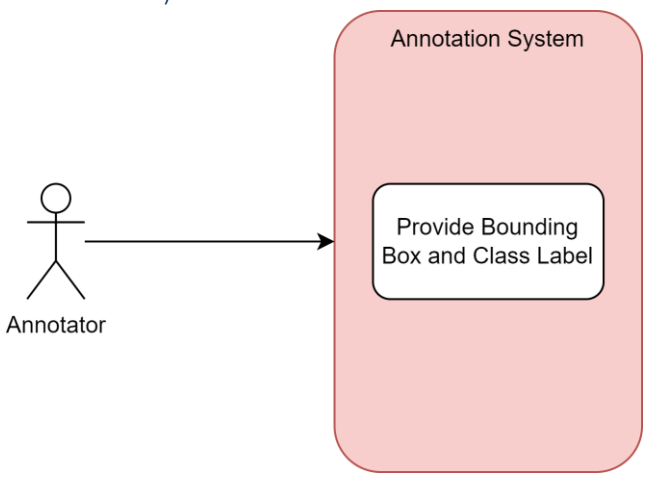
Configure Workflow Manager System (Mattia Di Donato)

In this phase we have these subtasks:

1. CRM Manager enters the parameters (7) of the new hospital into the system.
2. **APPLICATION**: notifies the inclusion of the new machine in the network.

Subtask	Actor	Cognitive Effort	Probability	Total Cost
Enters new hospital in the system	CRM Manager	1-Remember	100%	100%*1*1,08*7=7,56
Total Cost				7,56

Annotation System (Mattia Di Donato)



The Annotator is the only human actor in the system, he is an Otolaryngologist, and he oversees examining each photo and adding a bounding box and a classification label. We computed the cost for this actor by dividing the average annual salary in UK by £77000.

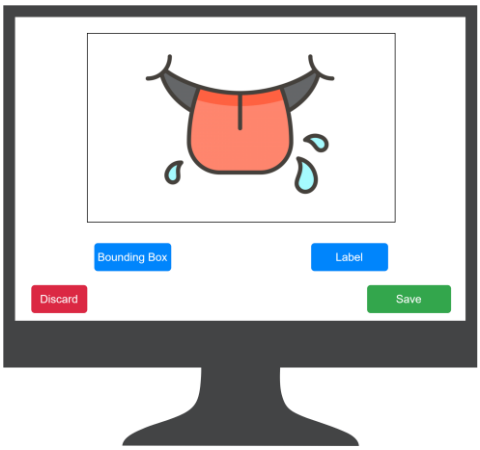
	Year	Cost
Annotator	£77000	77000 / 37000 = 2,08

Provide Bounding Box and Class Label (Mattia Di Donato)

In this phase we have these subtasks:

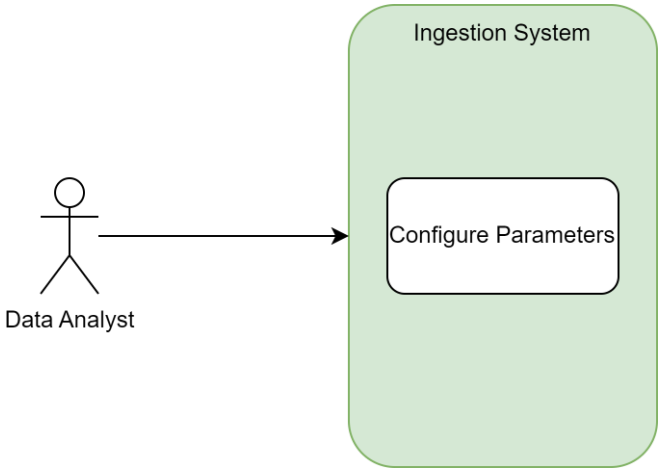
1. The annotator loads the oral lesion image from the data storage.
2. **Annotation System**: shows the oral lesion photo.
3. The annotator examines the image with his medical knowledge.
4. Adding bounding boxes.
5. The annotator classifies the image.
6. The annotator adds a class label.

7. Save classified image.
8. **Annotation System**: sent image and class label as samples.



Subtask	Actor	Cognitive Effort	Probability	Total Cost
Loads the oral lesion image	Otolaryngologist	1-Remember	100%	100%*1*2,08=2,08
Examines images	Otolaryngologist	4-Analyse	100%	100%*4*2,08=8,32
Adding bounding boxes	Otolaryngologist	3-Apply	100%	100%*3*2,08=6,24
Classifies the image	Otolaryngologist	4-Analyse	100%	100%*4*2,08=8,32
Add class label	Otolaryngologist	2-Understand	100%	100%*2*2,08=4,16
Save classified image	Otolaryngologist	1-Remember	100%	100%*1*2,08=2,08
Total Cost				31,2

Ingestion System (Mattia Di Donato)



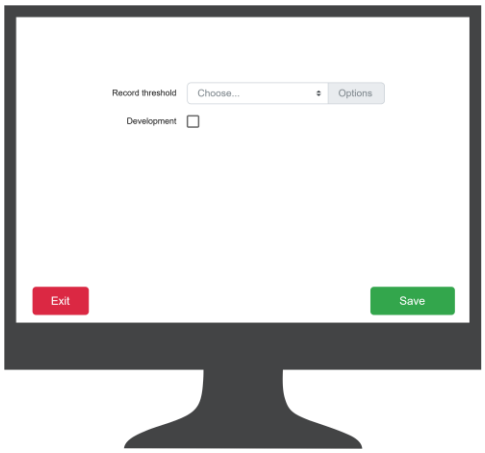
The data analyst is the only human actor involved in the ingestion phase. The data analyst is the only human actor involved in the ingestion phase. Here his task is to set the parameters about the errors on records and the development mode parameter. We computed the cost for this actor by dividing the average annual salary in UK by £37000, as well as the lowest annual salary of all those taken into consideration.

	Year	Cost
Data Analyst	£37000	37000 /37000 = 1

Configure Parameters – Ingestion System *(Mattia Di Donato)*

In this phase we have these subtasks:

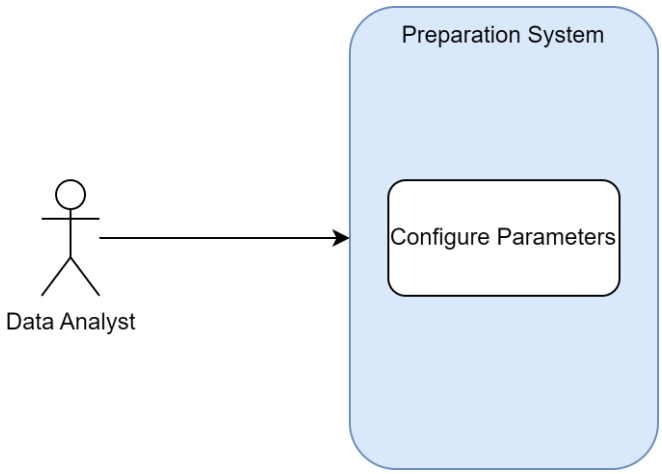
1. Data Analyst: set error threshold on record samples.
2. **Ingestion System**: show the entered values.
3. Data Analyst: set the development parameter.
4. **Ingestion System**: show the entered value.
5. Data Analyst: save the updates.
6. **Ingestion System**: adapts to the new parameters.



Subtask	Actor	Cognitive Effort	Probability	Total Cost
Set threshold on record samples	Data Analyst	3-Apply	100%	100%*3*1=3
Set development param	Data Analyst	1-Remember	100%	100%*1*1=1
Save updates	Data Analyst	1-Remember	100%	100%*1*1=1
Total Cost				5

Preparation System (Mattia Di Donato)

The data analyst is the only human actor involved in the preparation phase. Here his task is to set the brightness and contrast parameters on a sample image, the outlier criteria, the parameters for feature extraction, the total number of scans, the number of scans to be evaluated in the monitoring phase and the development mode parameter.



	Year	Cost
Data Analyst	£37000	37000 /37000 = 1

Configure Parameters – Preparation System *(Mattia Di Donato)*

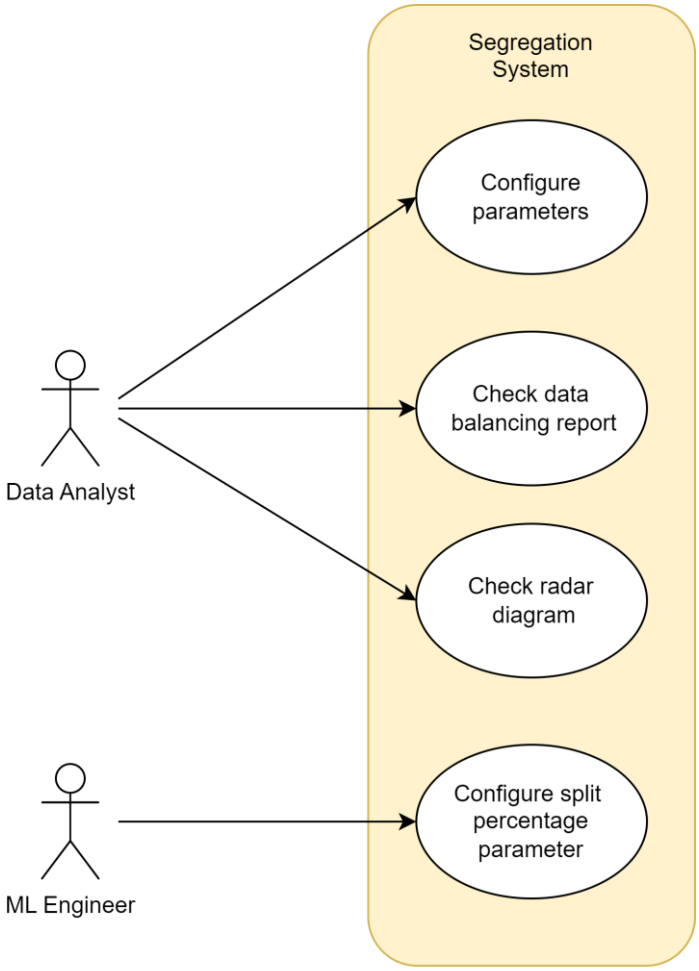
In this phase we have these subtasks:

1. Data Analyst: get a sample image of oral lesion.
2. **Preparation System**: show image into editor application.
3. Data Analyst: set brightness and contrast.
4. Data Analyst: save brightness and contrast parameters.
5. **Preparation System**: show the new brightness and contrast parameters applied to the image.
6. Data Analyst: set outlier criteria.
7. **Preparation System**: show the outlier criteria.
8. Data Analyst: set parameters for feature extraction.
9. **Preparation System**: show the parameters for future extraction.
10. Data Analyst: set number of max records.
11. **Preparation System**: show the entered value.
12. Data Analyst: set number of monitoring records.
13. **Preparation System**: show the entered value.
14. Data Analyst: set the development parameter.
15. **Preparation System**: show the entered value.
16. Data Analyst: save the updates.
17. **Preparation System**: adapts to the new parameters.

Subtask	**Actor**	**Cognitive Effort**	**Probability**	**Total Cost**
Get a sample image	Data Analyst	1-Remember	100%	100%*1*1=1
Set brightness and contrast	Data Analyst	4-Analyse	100%	100%*4*1=4
Save brightness and contrast params	Data Analyst	1-Remember	100%	100%*1*1=1
Set outlier criteria	Data Analyst	4-Analyse	100%	100%*4*1=4
Set params for features extraction	Data Analyst	4-Analyse	100%	100%*4*1=4
Set number of max scans	Data Analyst	2-Understand	100%	100%*2*1=2
Set number of check scans	Data Analyst	2- Understand	100%	100%*2*1=2
Set development param	Data Analyst	1-Remember	100%	100%*1*1=1
Save updates	Data Analyst	1-Remember	100%	100%*1*1=1

Total Cost				20

Segregation System (Matteo)



The actors involved in the segregation system are the Data Analyst and the ML engineer. We computed the normalized costs of this actors in this way.

	Year	Cost
Data Analyst	£37000	37000/37000 =1,00
ML Engineer	£64000	64000/37000 =1,73

Configure Parameters – Segregation System (Matteo)

1. Data Analyst clicks on “Set Ingestion parameters”.
2. **SYSTEM** show three input boxes related to ingestion threshold additional cases number and percentage variation tolerance.
3. Data Analyst choose the three parameters.
4. Data Analyst set the three parameters.
5. Data Analyst selects “Confirm”.
6. **SYSTEM** Store parameters.

Subtask	Actor	Cognitive Effort	Probability	Cost
Click on “Set Ingestion parameters”	Data Analyst	1 - Remember	100%	1 x 1 = 1
Choose parameters	Data Analyst	2 - Understand	100%	1 x 2 x 3 = 6
Set parameters	Data Analyst	1 - Remember	100%	1 x 1 x 3 = 3

Click on Confirm	Data Analyst	1 - Remember	100%	1 x 1 = 1
Total costs				11

Check data balancing report – Segregation System (Matteo)

1. Data Analyst request data balancing report.
2. **SYSTEM** display data balancing report and acceptation form.
3. Data Analyst calculate variance percentage for each class.
4. **IF** one class variance percentage is above the percentage variation tolerance
   4.1 Data Analyst declines data balancing
5. **ELSE**:
   5.1 Data Analyst approves data balancing.

Subtask	Actor	Cognitive Effort	Probability	Cost
Data balancing report request	Data Analyst	1 - Remember	100%	1 x 1 = 1
Calculate variance percentage for each class	Data Analyst	3 - Apply	100%	3 x 1 x 3 = 9
Compare classes variance with the percentage variation tolerance	Data Analyst	2 - Understand	100%	1 x 2 x 3 = 6
Declines data balancing	Data Analyst	1 - Remember	20%	1 x 1 x 0.2 = 0.2
Approves data balancing	Data Analyst	1 - Remember	80%	1 x 1 x 0.8 = 0.8
Total costs				17

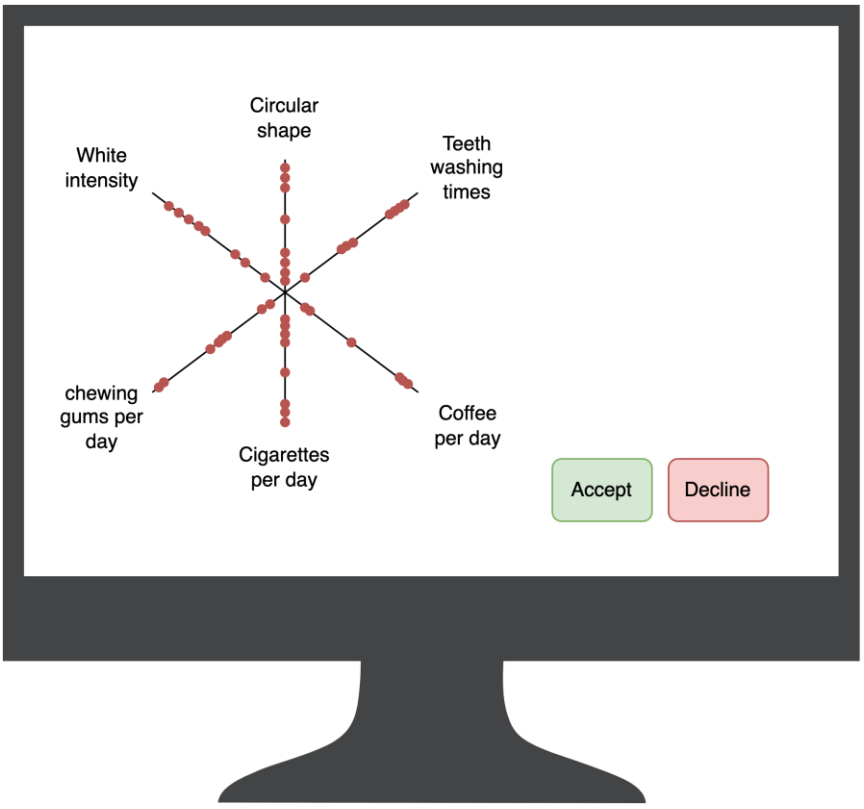
Rule to calculate variance percentage:

1. For each class:
   a. $\frac{\text{class\_instances\_mean} - \text{class\_instances}}{\text{class\_instances\_mean}} * 100$

Check radar diagram – Segregation System (Matteo)

1. Data Analyst requests radar diagram.
2. **SYSTEM** displays radar diagram report.
3. **IF** input coverage is not sufficient
   3.1 Data Analyst declines radar diagram report
4. **ELSE**
   4.1 Data Analyst approves radar diagram report.

Subtask	Actor	Cognitive Effort	Probability	Cost
Radar diagram report request	Data Analyst	1 - Remember	100%	1 x 1 = 1
Decide if input coverage is sufficient	Data Analyst	4 - Analyze	100%	1 x 4 = 4
Declines radar diagram report	Data Analyst	1 - Remember	10%	1 x 1 x 0.3= 0.1
Approves radar diagram report	Data Analyst	1 - Remember	90%	1 x 1 x 0.7 = 0.9
Total costs				6



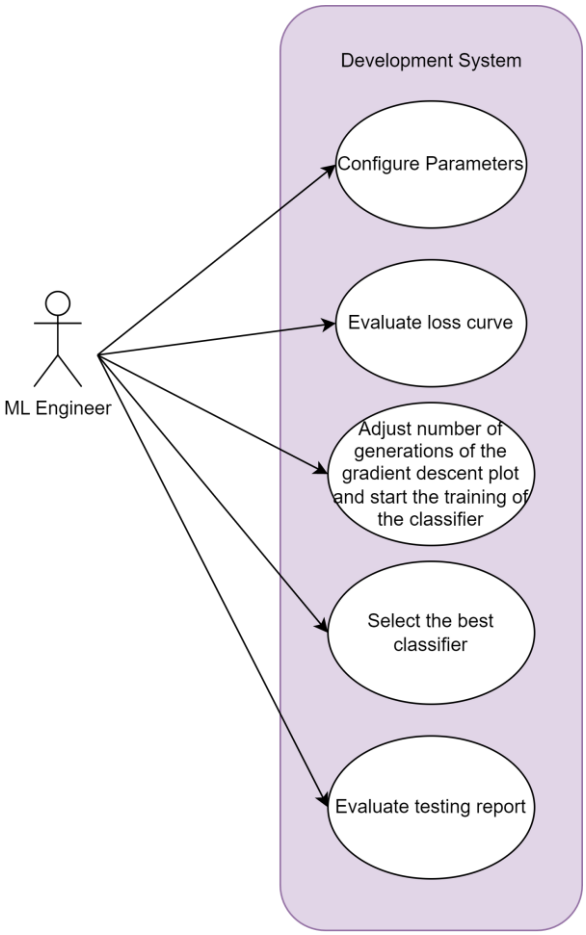
Configure split percentage parameter – Segregation System (Matteo)

1. ML Engineer clicks on “Set Segregation parameters”.
2. **SYSTEM** show three input boxes related to train, validation and test percentage.
3. ML Engineer decides split percentages.
4. ML Engineer set split percentage parameters.
5. ML Engineer selects “Confirm”.
6. **SYSTEM** Store parameters.

Subtask	Actor	Cognitive Effort	Probability	Cost
Click on “Set Segregation parameters”	ML Engineer	1 - Remember	100%	1,73 x 1 = 1,73

Decides split percentages	ML Engineer	2 - Understand	100%	1,73 x 2 x 3 = 10,38
Set split percentage parameters	ML Engineer	1 - Remember	100%	1,73 x 1 x 3 = 5,19
Click on Confirm	ML Engineer	1 - Remember	100%	1,73 x 1 = 1,73
Total costs				19,03

Development System (Giuseppe)



Set development parameters – Development system

1. ML Engineer clicks on “Set Parameters”.
2. FOR EACH hyperparameter
   2.1. **SYSTEM** show the hyperparameters (Number of neurons and number of layers) with the input fields to insert maximum and minimum values
   2.2. ML Engineer selects the minimum value of the hyperparameter
   2.3. ML Engineer selects the maximum value of the hyperparameter
3. **SYSTEM** show the input field to insert parameters values
   3.1. ML Engineer selects the validation tolerance value
   3.2. ML Engineer selects the testing tolerance value
4. ML Engineer selects “Confirm”

Subtask	**Actor**	**Cognitive Effort**	**Occurrences**	**Cost**
Click on “Set Parameters”	ML Engineer	1 - Remember	1	1*1*1,73 = 1,73
Select the minimum value of the hyperparameter	ML Engineer	2 - understand	2	2*2*1.73 = 6,92

Select the maximum value of the hyperparameter	ML Engineer	2 - understand	2	2*2*1,73 = 6,92
Select the validation tolerance value	ML Engineer	2 - understand	1	2*1*1,73 = 3,46
Select the testing tolerance value	ML Engineer	2 - understand	1	2*1*1,73 = 3,46
Click on "Confirm"	ML Engineer	1 - Remember	1	1*1*1,73 = 1.73
Total cost				24,22 £

Evaluate loss curve.

1. ML Engineer clicks on "Show training error plot".
2. SYSTEM plots the loss curve.
3. **IF** the curve flattens at the end and it is flat for less than half of the graph
   3.1. ML Engineer clicks on "Num of generations ok"
4. **ELSE**
   4.1. ML Engineer clicks on "Num of generations not ok"

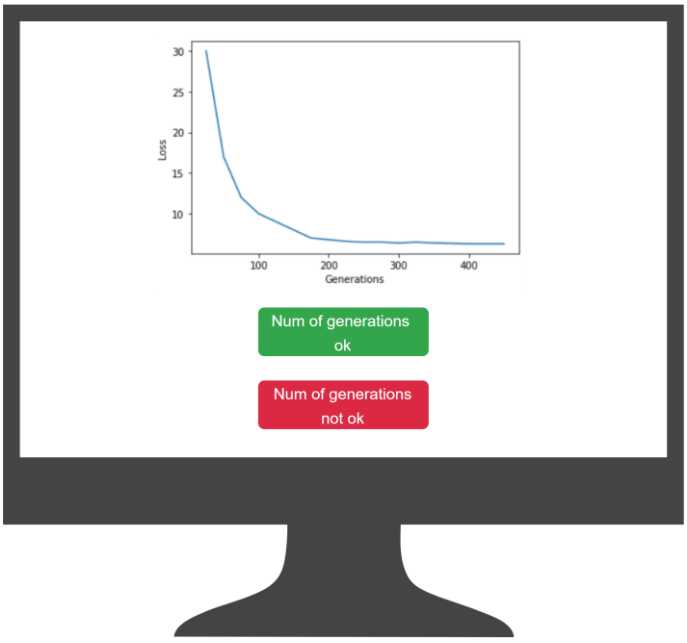


Figure 1 Num of generations not ok

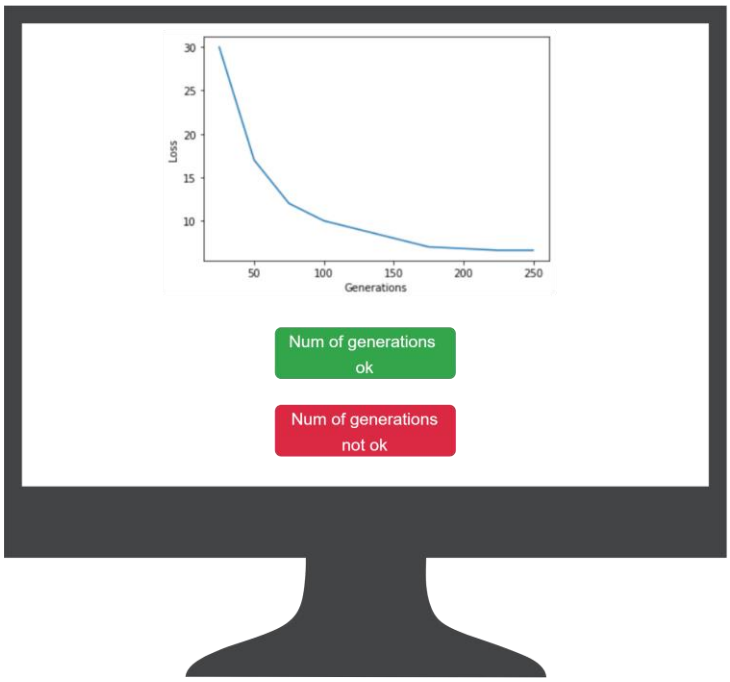


Figure 2 Num of generations ok

Subtask	**Actor**	**Cognitive Effort**	**Occurrences**	**Cost**
Click on "Show training error plot"	ML Engineer	1 - Remember	100%	1*1*1,73 = 1,73
Evaluate if the curve is flat for less than half of the graph	ML Engineer	3 - Apply	100%	3*1*1,73 = 5,19
Click on "Num of generations not ok"	ML Engineer	1 - Remember	30%	1*0,7*1,73 = 0,52
Click on "Num of generations ok"	ML Engineer	1 - Remember	70%	1*0,7*1,73 = 1,21
Total cost				8,65 £

Adjust number of generations of the gradient descent plot and start the training of the classifier.

1. SYSTEM shows the input field for inserting the number of generations
2. ML Engineer decides the number of generations
3. ML Engineer clicks on "Train the classifier".

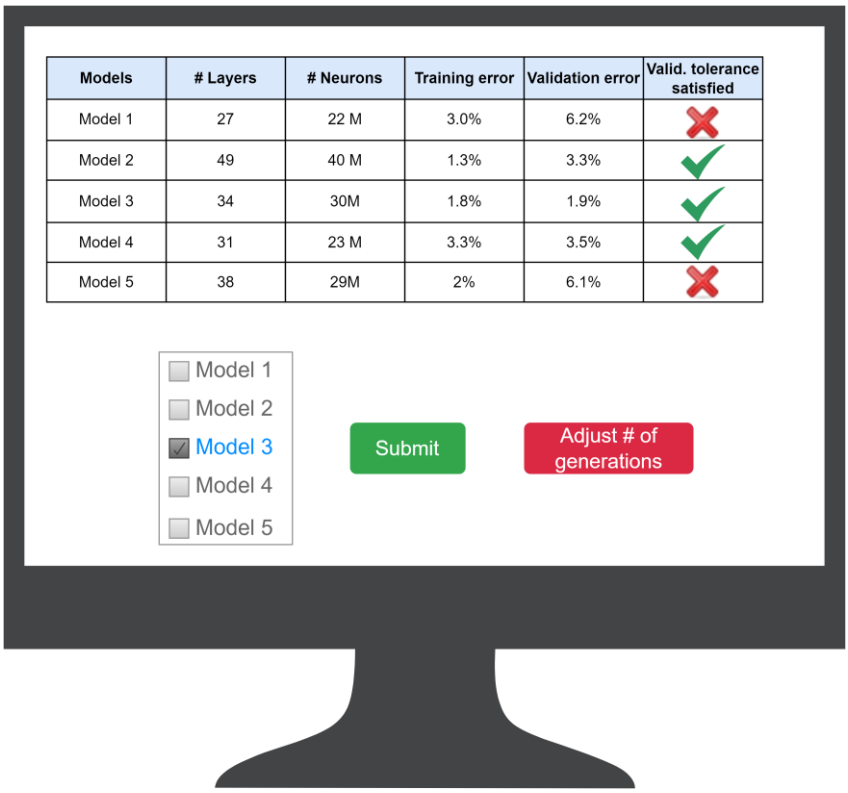
Subtask	**Actor**	**Cognitive Effort**	**Occurrences**	**Cost**
Decide the number of generations	ML Engineer	3 - Apply	100%	3*1*1,73 = 5,19
Click on "Start the training"	ML Engineer	1 - Remember	100%	1*1*1,73 = 1,73
Total cost				6,92 £

Rule to decide the number of generations:

1. If there is the need to adjust the number of generations because the loss curve is not acceptable:
    a. if the loss curve is not flat at the end of the iterations
        i. enlarge by one third the previous number of generations
    b. if the loss curve is flat for more than the half of the graph (overfitting)
        i. reduce by one third the number of generations
2. if there is the need to adjust the number of generations because the difference between validation and training error is under the validation tolerance (overfitting)
    a. reduce by one third the number of generations

Select the best classifier.

1. ML Engineer clicks on “Generate validation report”
2. **SYSTEM** show the validation report with the best 5 networks
3. ML Engineer selects the best classifier
4. ML Engineer compares the difference between validation and training error with the validation tolerance.
5. ML Engineer: **IF** the difference between validation and training error is under the validation tolerance
   5.1. ML Engineer clicks on the box of the selected classifier
   5.2. ML Engineer clicks on “Submit”
6. **ELSE**: ML Engineer clicks on “Adjust # of iterations”



Subtask	Actor	Cognitive Effort	Occurrences	Cost
Click on “Generate validation report”	ML Engineer	1 - Remember	100%	1*1*1,73 = 1,73
Select the best classifier	ML Engineer	3 - Apply	100%	3*1*1,73 = 5,19
compares the difference between validation and training error with the validation tolerance	ML Engineer	2 - understand	100%	2*1*1,73 = 3,46
Click on the box of the selected classifier	ML Engineer	1 - Remember	80%	1*0,8*1,73 = 1,38
Click on “Submit”	ML Engineer	1 - Remember	80%	1*0,8*1,73 = 1,38
Click on “Adjust # of generations”	ML Engineer	1 - Remember	20%	1*0,2*1,73 = 0,35
Total cost				13,49 £

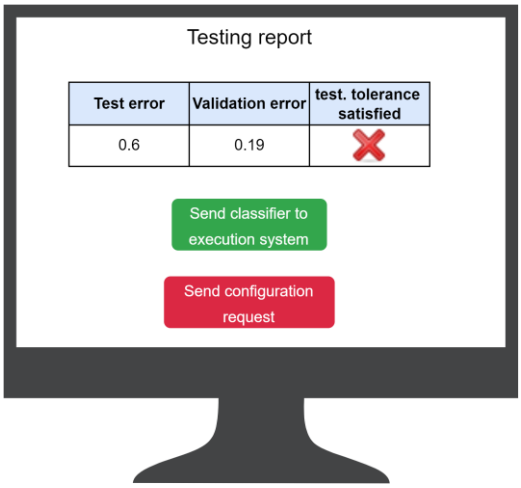
Rules to select the best classifier:

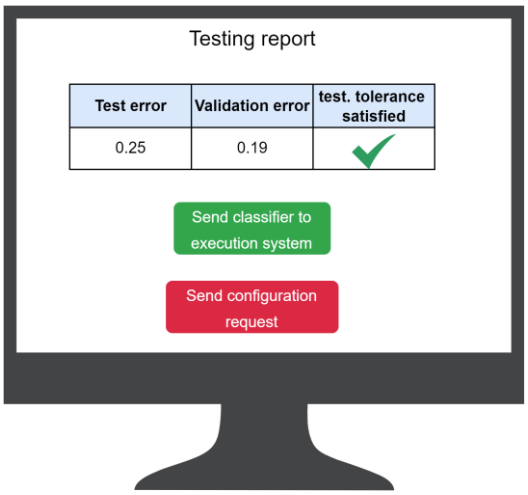
1. Select the model with the lowest validation error and not already evaluated.

2. **if** the difference between validation and training error is over the validation tolerance:
    a. if not all 5 classifiers have already been evaluated
        i. return to 1
    b. else
        i. select this current classifier as the best one.
3. **Else**
    a. consider other classifiers with a similar validation error to the current selected one and a difference between validation and training error under the validation errors if any.
    b. Select the least complex classifier as the best one.

Evaluate testing report.

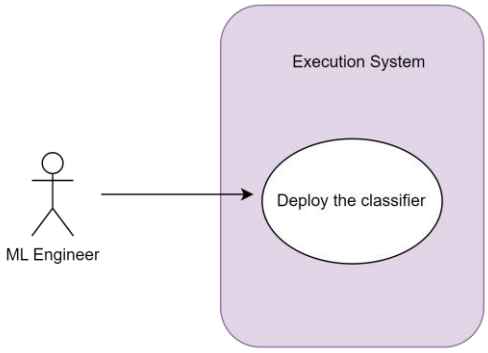
1. ML Engineer clicks on “Generate testing report”
2. SYSTEM displays the testing report
3. ML Engineer compares the difference between validation and testing error with the testing tolerance.
4. **IF** the difference between validation and testing error is under the testing tolerance
    4.1. ML Engineer clicks on “Send classifier to the execution system”.
5. **ELSE**
    5.1. ML Engineer clicks on “Send configuration request” for the annotation of additional cases.





Subtask	Actor	Cognitive Effort	Occurrences	Cost
Click on “Generate testing report”	ML Engineer	1 - Remember	100%	1*1*1,73 = 1,73
compares the difference between validation and testing error with the testing tolerance	ML Engineer	2 - understand	100%	2*1*2,73 = 5,46
clicks on “Send classifier to the execution system”	ML Engineer	1 - Remember	90%	1*0,9*1,73 = 1,56
clicks on “Send a configuration request for the annotation of additional cases”	ML Engineer	1 - Remember	10%	1*0,1*1,73 = 0,17
Total cost				8,92 £

Execution system (Giuseppe)

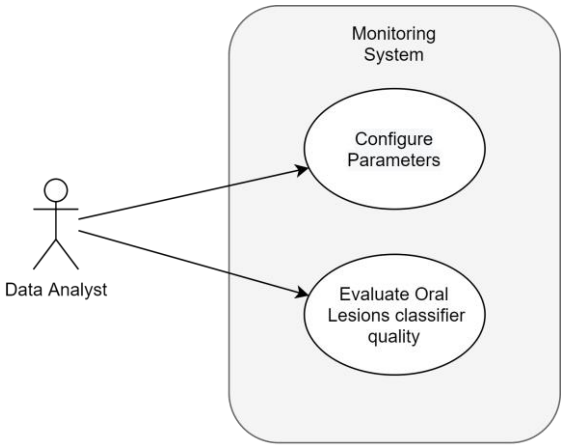


Deploy oral lesions classifier

1. SYSTEM shows the interface for the deployment
2. ML Engineer: clicks on "deploy the classifier"

Subtask	**Actor**	**Cognitive Effort**	**Occurrences**	**Cost**
Clicks on "Deploy classifier"	ML Engineer	1 - Remember	100%	1*1*1,73 = 1,73
Total cost				1,73£

Monitoring System (Salvatore Arancio Febbo)



	Year	*Cost*
Data Analyst	£37000	37000/37000=1

The only human actor involved is the Data Analyst. We computed the cost for this actor by dividing the average annual salary in UK by £37000, as well as the lowest annual salary of all those taken into consideration.

Set monitoring parameter- Monitoring System (Salvatore Arancio Febbo)

1. Data Analyst: set total records to monitor the classifier
2. Data Analyst: set the number of labeling errors
3. Data Analyst: set the number of consecutive labeling errors
4. Data Analyst: set workflow process address
5. System: Store parameters.

Subtask	**Actor**	**Cognitive Effort**	**Probability**	**Cost**
Set total records to monitor the classifier	Data Analyst	2 - understand	100%	1 x 1 x 2 = 2
Set the number of labeling errors	Data Analyst	2 - understand	100%	1 x 1 x 2 = 2

Set the number of consecutive labeling errors	Data Analyst	2 - understand	100%	1 x 1 x 2 = 2
Set workflow process address	Data Analyst	1-Remember	100%	1 x 1 x 1 = 1
Total cost				7.00

Justify Cognitive Levels:

- set total records to monitor the classifier (2 - understand) because there is a statistical categorization.
- set the number of labeling errors and set the number of consecutive labeling errors (2 - understand) because there are satisfied and exceeded categories.
- set workflow process address (1-Remember) because this step involves recalling or remembering previously learned information.

Evaluate oral lesions session classifier quality - Monitoring System (Salvatore Arancio Febbo)

1. Data analyst: Search oral lesions sessions classifier and expert's report
2. System: Display oral lesions sessions expert's report
3. System: Display classifier quality form
4. Data Analyst: Analyze oral lesions sessions quality report
5. IF Max. number of labeling errors reached
   5.1. Data Analyst: Notify classifier's low quality
6. ELSE
   6.1. Data Analyst: Approve classifier quality

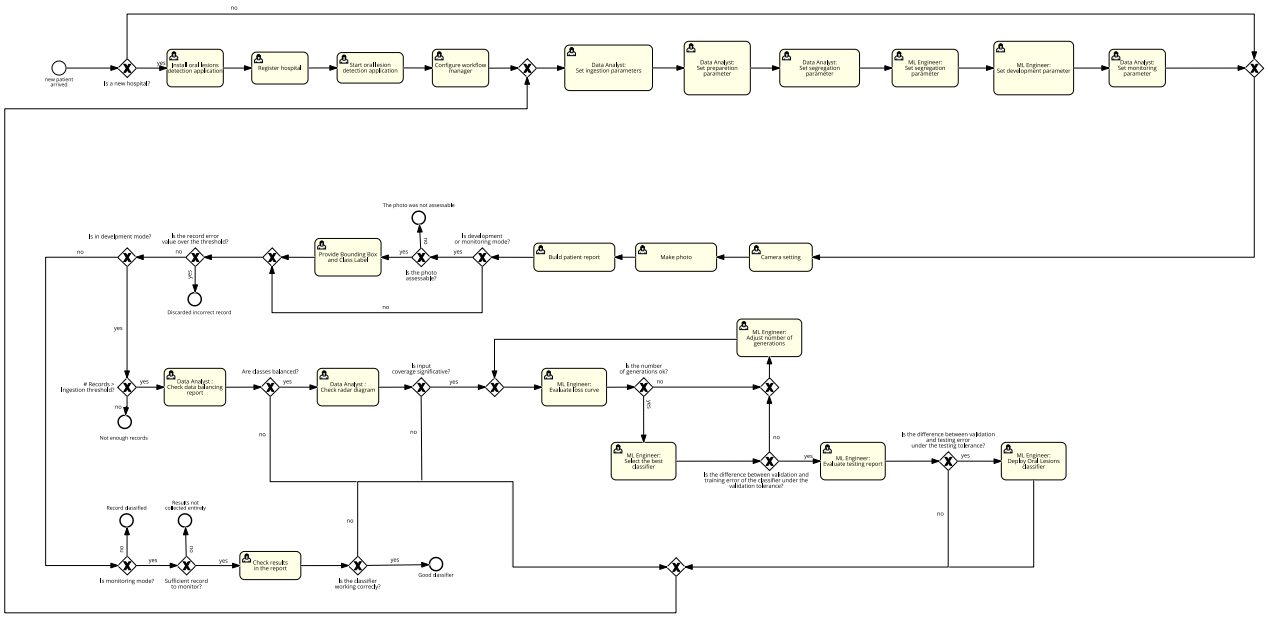
Subtask	**Actor**	**Cognitive Effort**	**Probability**	**Cost**
Search oral lesions sessions classifier and expert's report	Data analyst	1 - Remember	100%	1x1x1 = 1
Analyze oral lesions sessions quality report	Data analyst	2 - understand	100%	1x1x2= 2
Notify classifier's low quality	Data analyst	1 - Remember	30%	1x1x0.3 = 0.3
Approve classifier quality	Data analyst	1 - Remember	70%	1x1x0.7 = 0.7
Total cost				4

Justify Cognitive Levels:

- Search oral lesions sessions classifier and expert's report, notify classifier's low quality and approve classifier quality (1-Remember) because this step involves recalling or remembering previously learned information.
- Analyze oral lesions sessions quality report (2 - understand) because there are the categories.

BIMP SIMULATION

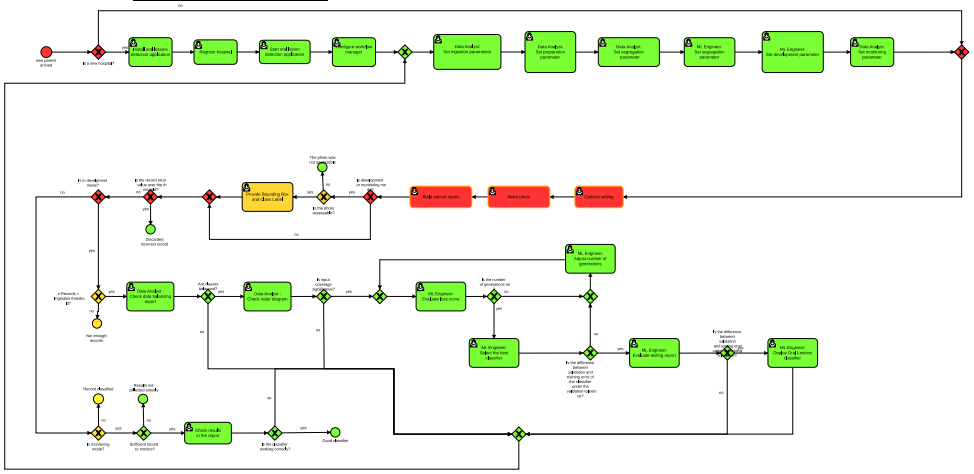
AS-IS (Salvatore, Matteo)



Gateway	Yes (%)	No (%)	Description
Is a new hospital?	0.02	99.98	Only 2 tokens over the total 10000 will be for hospital registration and initial configuration
Is development or monitoring mode?	55	45	1000 records for each hospital, 500 for development 450 for execution and 50 for monitoring
Is the photo assessable?	95	5	In this case only the 5% of the photos are not assessable.
Is the record error value over the threshold?	0.1	99.9	In this case only the 0.1% of the records are discarded because they contain too many errors on essential information.
Is in development mode?	50	50	Records for the development mode are 500 over 1000
# Records > Ingestion threshold?	1	99	If the ingestion threshold is set to 500 only 1 token should pass the gateway in the yes direction (0.2%), but we should also consider the additional cases caused by possible reconfiguration, so we decide for 1% split.
Are classes balanced?	80	20	We suppose that 20% of times classes are not balanced
Is input coverage significative?	90	10	We suppose that only 10% of times the feature covering is not significative
Is the number of generations, ok?	70	30	We suppose that 3 cycles are sufficient to correct the number of generations. So, to have some tokens which cycles 3 times we set the probability of 30% on the no branch. (0.3*0.3*0.3 *100 = 2.7 ~ 3 tokens over 100)
Is the difference between validation and training error of the classifier under the validation tolerance?	80	20	We suppose that 80% of times training and validation errors are similar enough (difference under the validation tolerance)
Is the difference between validation and testing error under the testing tolerance?	90	10	We suppose that 90% of times testing and validation errors are similar enough (difference under the testing tolerance)
Is monitoring mode?	10	90	About 50% of the total tokens will go through here, only 1/10th will be considered for monitoring, the rest will go into execution.
Sufficient record to monitor?	2	98	Waiting 50 tokens, then I'll put an access probability of 1/50th.
Is the classifier working correctly?	90	10	Assuming that 90% of the time the classifier gives good results, the remaining percentage of the time the model will have to be retrained

Heatmap

Heatmap based on Counts

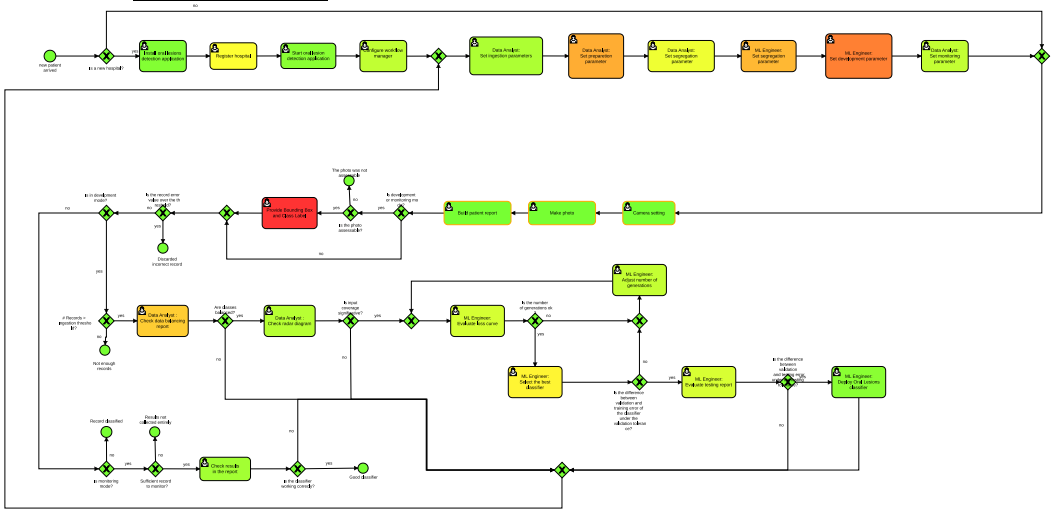


Legend

Color	Value
	0
	1118
	2236
	3355
	4473
	5591
	6709
	7828
	8946
	10064

Heatmap

Heatmap based on Durations



Legend

Color	Value
	0 s
	3 s
	7 s
	10 s
	14 s
	17 s
	21 s
	24 s
	28 s
	31 s

AS-IS Statistics

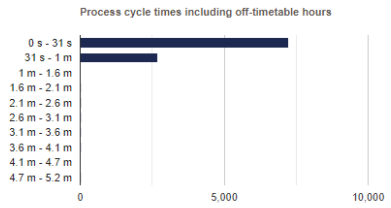
General information

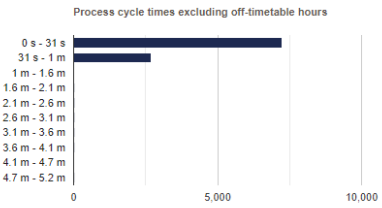
Completed process instances 10000

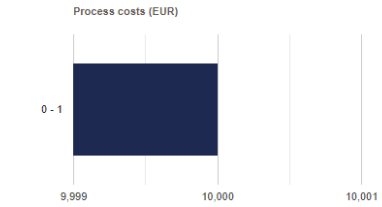
Total cost 0 EUR

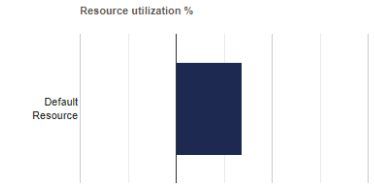
Total simulation time 41.4 weeks

Charts









Scenario Statistics

	Minimum	Maximum	Average
Process instance cycle times including off-timetable hours	0 seconds	5 minutes	17.2 seconds
Process instance cycle times excluding off-timetable hours	0 seconds	5 minutes	17.2 seconds
Process instance costs	0 EUR	0 EUR	0 EUR

Activity Durations, Costs, Waiting times, Deviations from Thresholds

Name		Waiting time			Duration			Duration over threshold			Cost			Cost over threshold		
	Count	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
Check results
in the report	8	0 s	0 s	0 s	3.7 s	4 s	4.4 s	0 s	0 s	0 s	0	0	0	0	0	0
Configure workflow manager
1	0 s	0 s	0 s	6.8 s	6.8 s	6.8 s	0 s	0 s	0 s	0	0	0	0	0	0
Data Analyst :
Check data balancing report	62	0 s	0 s	0 s	15.4 s	17 s	18.7 s	0 s	0 s	0 s	0	0	0	0	0	0
Data Analyst :
Check radar diagram	50	0 s	0 s	0 s	5.4 s	5.9 s	6.6 s	0 s	0 s	0 s	0	0	0	0	0	0
Data Analyst:
Set ingestion parameters
65	0 s	0 s	0 s	4.5 s	5 s	5.5 s	0 s	0 s	0 s	0	0	0	0	0	0
Data Analyst:
Set monitoring parameter	65	0 s	0 s	0 s	6.4 s	7 s	7.7 s	0 s	0 s	0 s	0	0	0	0	0	0
Data Analyst:
Set preparetion parameter	65	0 s	0 s	0 s	18 s	20 s	22 s	0 s	0 s	0 s	0	0	0	0	0	0
Data Analyst:
Set segregation
parameter	65	0 s	0 s	0 s	9.9 s	11.1 s	12.1 s	0 s	0 s	0 s	0	0	0	0	0	0
Install oral lesions detection application	1	0 s	0 s	0 s	1.1 s	1.1 s	1.1 s	0 s	0 s	0 s	0	0	0	0	0	0
ML Engineer:
Adjust number of generations	24	0 s	0 s	0 s	6.4 s	7.1 s	7.6 s	0 s	0 s	0 s	0	0	0	0	0	0
ML Engineer:
Deploy Oral Lesions classifier	39	0 s	0 s	0 s	1.6 s	1.7 s	1.9 s	0 s	0 s	0 s	0	0	0	0	0	0
ML Engineer:
Evaluate loss curve	64	0 s	0 s	0 s	7.8 s	8.6 s	9.5 s	0 s	0 s	0 s	0	0	0	0	0	0
ML Engineer:
Evaluate testing report	40	0 s	0 s	0 s	8 s	9 s	9.7 s	0 s	0 s	0 s	0	0	0	0	0	0
ML Engineer:
Select the best classifier	45	0 s	0 s	0 s	12.1 s	13.4 s	14.4 s	0 s	0 s	0 s	0	0	0	0	0	0
ML Engineer:
Set development parameter	65	0 s	0 s	0 s	21.9 s	24.1 s	26.6 s	0 s	0 s	0 s	0	0	0	0	0	0
ML Engineer:
Set segregation parameter	65	0 s	0 s	0 s	17.2 s	19 s	20.9 s	0 s	0 s	0 s	0	0	0	0	0	0
Provide Bounding Box and Class Label	5255	0 s	0 s	0 s	28.1 s	31.2 s	34.3 s	0 s	0 s	0 s	0	0	0	0	0	0
Register hospital	1	0 s	0 s	0 s	12.4 s	12.4 s	12.4 s	0 s	0 s	0 s	0	0	0	0	0	0
Start oral lesion detection application	1	0 s	0 s	0 s	1.8 s	1.8 s	1.8 s	0 s	0 s	0 s	0	0	0	0	0	0

Differences between AS-IS and TO-BE

Handoff level (Salvatore Arancio Febbo)

In the current system (AS-IS), if the classes are not balanced before training, the radar diagram may not have enough features and if the difference between the test and validation errors is below the testing tolerance, this indicates that the data is not goof and a request for new data is sent to the workflow manager.

In the updated system (TO-BE), if the classes are not balanced before training, the radar diagram may not have enough features and if the difference between the test and validation errors is below the testing tolerance, this indicates that the data is not good. To address this issue, a new task service has been added which retrieves new records from other similar hospitals rather than requesting new data from the workflow manager.

Service level (Giuseppe)

At service level we decided to skip hyperparameters tuning via grid search.

We assumed that the optimal hyperparameters are chosen from a classifier that the company has already developed for a similar hospital (similar in terms of hospital dimension and percentage of foreign patients).

Consequently, the task "select the best classifier" is eliminated, since we no longer have different classifiers for each different combination of hyperparameters. It is replaced by the task “Evaluate validation report”, where the ML Engineer has just to evaluate whether the difference between the training and validation error of the unique classifier is under the validation tolerance or not.

Additionally, the cost of the task "Set development parameters" is reduced because the minimum and maximum values for the range of hyperparameters no longer have to be set in the configuration of the development system.

Below is the cost for the task “Evaluate validation report” and “Set development parameters” in the to-be.

Evaluate validation report

Subtask	**Actor**	**Cognitive Effort**	**Occurrences**	**Cost**
Click on “Generate validation report”	ML Engineer	1 - Remember	100%	1*1*1,73 = 1,73
compares the difference between validation and training error with the validation tolerance	ML Engineer	2 - understand	100%	2*1*1,73 = 3,46
Click on “Submit”	ML Engineer	1 - Remember	80%	1*0,8*1,73 = 1,38
Click on “Adjust # of generations”	ML Engineer	1 - Remember	20%	1*0,2*1,73 = 0,35
Total cost				6,92 £

So the cost of the previous task of 13,49 is replaced by the cost of the new task which is 6,92

Set development parameters

Subtask	**Actor**	**Cognitive Effort**	**Occurrences**	**Cost**
Click on “Set Parameters”	ML Engineer	1 - Remember	1	1*1*1,73 = 1,73
Select the validation tolerance value	ML Engineer	2 - understand	1	2*1*1,73 = 3,46
Select the testing tolerance value	ML Engineer	2 - understand	1	2*1*1,73 = 3,46
Select the optimal value of hyperparameters	ML Engineer	2 - understand	2	2*2*1,73 = 6,92
Click on “Confirm”	ML Engineer	1 - Remember	1	1*1*1,73 = 1.73

Total cost				17,3 £

The cost of the task is reduced from 24,22 to 17,3.

Task level (Matteo)

To improve the performance of our factory we decided do make the balancing report more automatized because it has a cost of 17 and as shown in the simulation is repeated more or less 62 times.

The new balancing report provides a table with all the variance percentage and an intuitive way to understand if the percentage is over or under the threshold. In this way the subtask with cognitive level of 3 is removed and the cost of comparison task is reduced from 6 to 2.

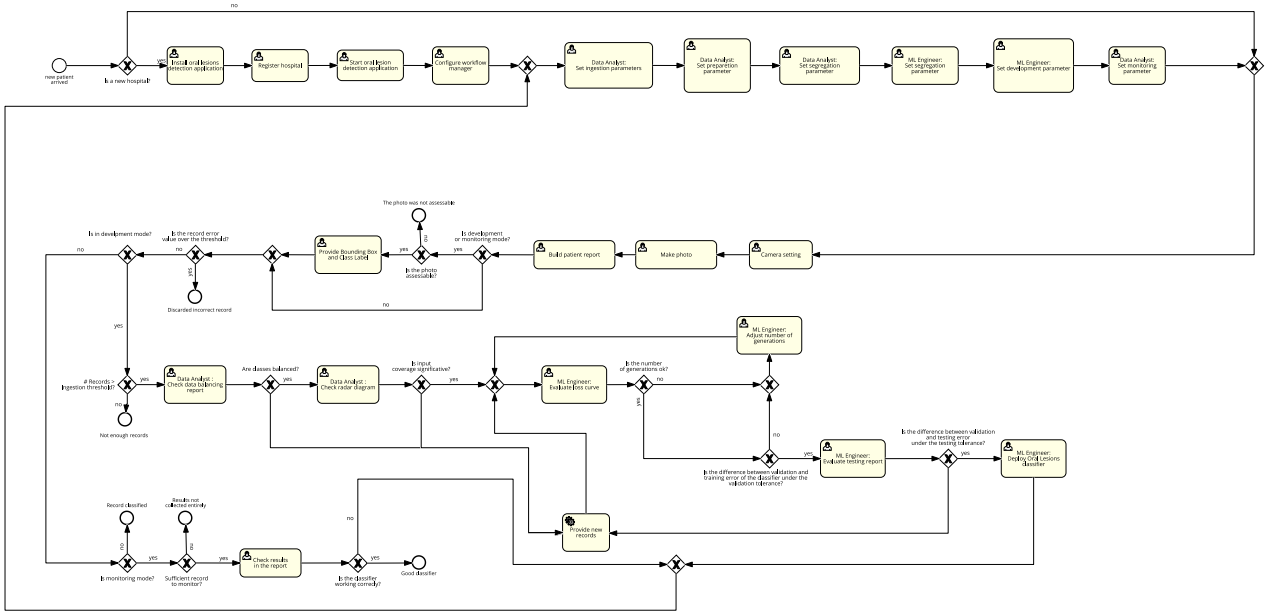
In this way we moved from 62 * 17 = 1054 to 44 * 4 = 176 reducing 6 times the total cost of this task.

Check data balancing report – Task level improvement.

1. Data Analyst request data balancing report.
2. **SYSTEM** display data balancing report and acceptation form.
3. **IF** one class variance percentage is above the percentage variation tolerance
   3.1 Data Analyst declines data balancing
4. **ELSE**:
   4.1 Data Analyst approves data balancing.

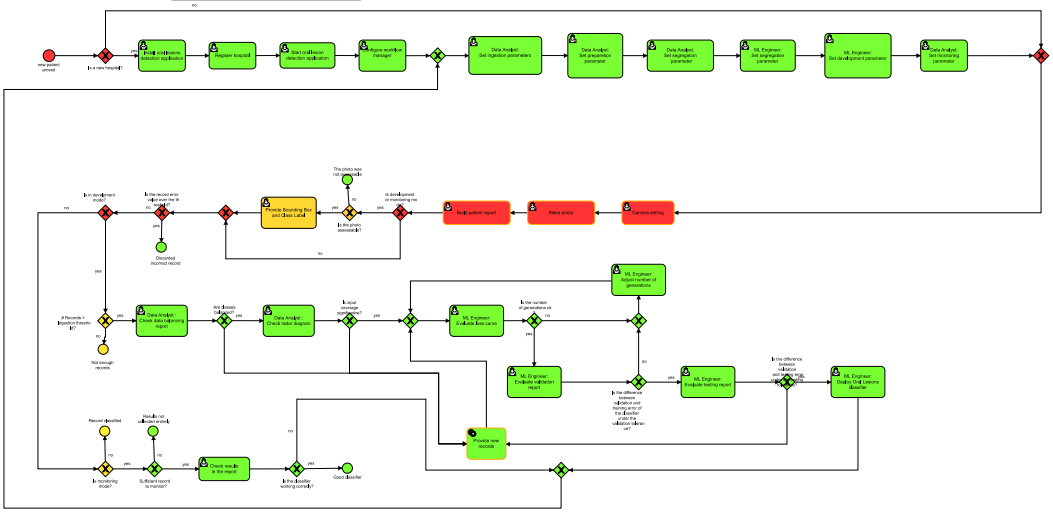
Subtask	**Actor**	**Cognitive Effort**	**Probability**	**Cost**
Data balancing report request	Data Analyst	1 - Remember	100%	1 x 1 = 1
Compare classes variance with the percentage variation tolerance	Data Analyst	2 - Understand	100%	1 x 2 = 2
Declines data balancing	Data Analyst	1 - Remember	20%	1 x 1 x 0.2 = 0.2
Approves data balancing	Data Analyst	1 - Remember	80%	1 x 1 x 0.8 = 0.8
Total costs				4

TO-BE (Giuseppe Martino, Mattia Di Donato)



Heatmap

Heatmap based on Counts

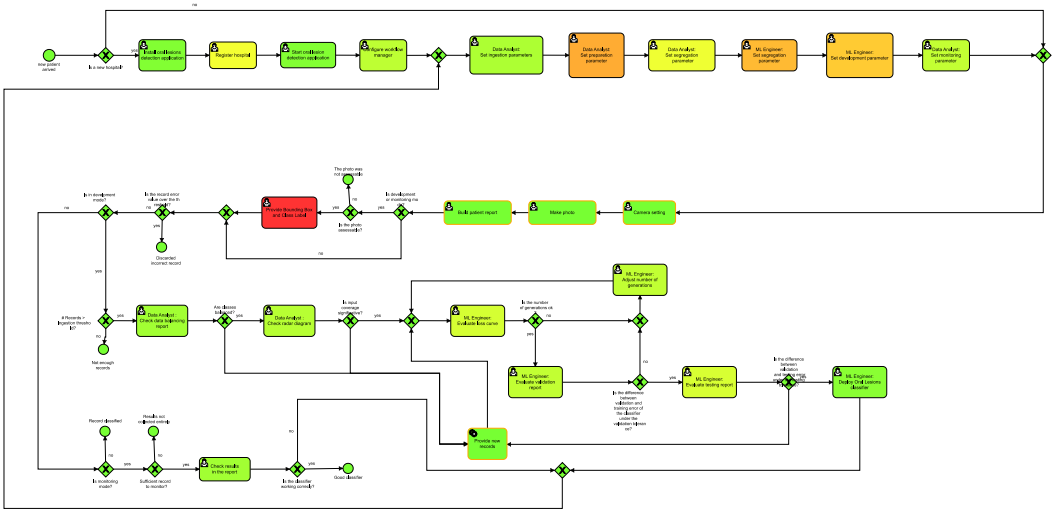


Legend

Color	Value
	0
	1117
	2234
	3351
	4468
	5586
	6703
	7820
	8937
	10054

Heatmap

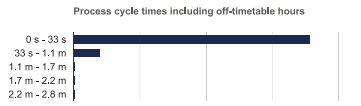
Heatmap based on Durations

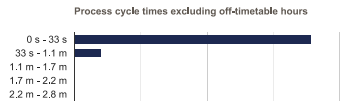


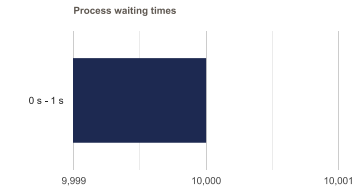
Legend

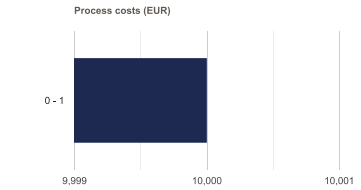
Color	Value
	0 s
	3 s
	7 s
	10 s
	14 s
	17 s
	21 s
	24 s
	28 s
	31 s

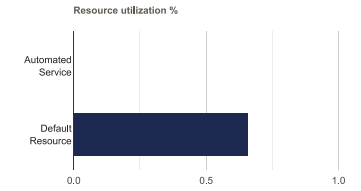
TO-BE Statistics











Scenario Statistics

	Minimum	Maximum	Average
Process instance cycle times including off-timetable hours	0 seconds	5.5 minutes	16.7 seconds
Process instance cycle times excluding off-timetable hours	0 seconds	5.5 minutes	16.7 seconds
Process instance costs	0 EUR	0 EUR	0 EUR

Activity Durations, Costs, Waiting times, Deviations from Thresholds

Name		Waiting time			Duration			Duration over threshold			Cost			Cost over threshold		
	Count	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
Check results
in the report	14	0 s	0 s	0 s	3.6 s	4 s	4.3 s	0 s	0 s	0 s	0	0	0	0	0	0
Configure workflow manager
2	0 s	0 s	0 s	6.1 s	6.6 s	7.1 s	0 s	0 s	0 s	0	0	0	0	0	0
Data Analyst :
Check data balancing report	44	0 s	0 s	0 s	3.6 s	4 s	4.4 s	0 s	0 s	0 s	0	0	0	0	0	0
Data Analyst :
Check radar diagram	42	0 s	0 s	0 s	5.4 s	5.9 s	6.6 s	0 s	0 s	0 s	0	0	0	0	0	0
Data Analyst:
Set ingestion parameters
47	0 s	0 s	0 s	4.5 s	5 s	5.4 s	0 s	0 s	0 s	0	0	0	0	0	0
Data Analyst:
Set monitoring parameter	47	0 s	0 s	0 s	6.3 s	7 s	7.7 s	0 s	0 s	0 s	0	0	0	0	0	0
Data Analyst:
Set preparetion parameter	47	0 s	0 s	0 s	18.2 s	20.1 s	21.9 s	0 s	0 s	0 s	0	0	0	0	0	0
Data Analyst:
Set segregation
parameter	47	0 s	0 s	0 s	10 s	11 s	11.9 s	0 s	0 s	0 s	0	0	0	0	0	0
Install oral lesions detection application	2	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	0	0	0	0	0	0
ML Engineer:
Adjust number of generations	45	0 s	0 s	0 s	6.3 s	6.8 s	7.6 s	0 s	0 s	0 s	0	0	0	0	0	0
ML Engineer:
Deploy Oral Lesions classifier	44	0 s	0 s	0 s	1.6 s	1.7 s	1.9 s	0 s	0 s	0 s	0	0	0	0	0	0
ML Engineer:
Evaluate loss curve	97	0 s	0 s	0 s	7.8 s	8.6 s	9.5 s	0 s	0 s	0 s	0	0	0	0	0	0
ML Engineer:
Evaluate testing report	52	0 s	0 s	0 s	8 s	8.9 s	9.7 s	0 s	0 s	0 s	0	0	0	0	0	0
ML Engineer:
Evaluate validation report	68	0 s	0 s	0 s	6.3 s	6.9 s	7.6 s	0 s	0 s	0 s	0	0	0	0	0	0
ML Engineer:
Set development parameter	47	0 s	0 s	0 s	15.6 s	17.2 s	18.8 s	0 s	0 s	0 s	0	0	0	0	0	0
ML Engineer:
Set segregation parameter	47	0 s	0 s	0 s	17.2 s	18.8 s	20.9 s	0 s	0 s	0 s	0	0	0	0	0	0
Provide Bounding Box and Class Label	5160	0 s	0 s	0 s	28.1 s	31.2 s	34.3 s	0 s	0 s	0 s	0	0	0	0	0	0
Register hospital	2	0 s	0 s	0 s	10.9 s	11.5 s	12.2 s	0 s	0 s	0 s	0	0	0	0	0	0
Start oral lesion detection application	2	0 s	0 s	0 s	1.1 s	1.1 s	1.1 s	0 s	0 s	0 s	0	0	0	0	0	0

Cost difference between AS-IS and TO-BE (Mattia Di Donato)

Task	AS IS	AS IS duration	AS IS total duration	TO BE	TO BE duration	TO BE total duration	IMPROVEMENT duration	IMPROVEMENT %
	Count	Avg Duration (s)	Count * Avg duration (s)	Count	Avg Duration (s)	Count * Avg duration (s)	As Is - To Be	%
Check results in the report	8.00	4.00	32.00	14.00	4.00	56.00	0.00	0.00
Configure workflow manager	1.00	6.80	6.80	2.00	6.60	13.20	0.20	2.94
Data Analyst: Check data balancing report	62.00	17.00	1054.00	44.00	4.00	176.00	13.00	76.47
Data Analyst: Check radar diagram	50.00	5.90	295.00	42.00	5.90	247.80	0.00	0.00
Data Analyst: Set ingestion parameters	65.00	5.00	325.00	47.00	5.00	235.00	0.00	0.00
Data Analyst: Set monitoring parameter	65.00	7.00	455.00	47.00	7.00	329.00	0.00	0.00
Data Analyst: Set preparetion parameter	65.00	20.00	1300.00	47.00	20.10	944.70	-0.10	-0.50
Data Analyst: Set segregation parameter	65.00	11.10	721.50	47.00	11.00	517.00	0.10	0.90
Install oral lesions detection application	1.00	1.10	1.10	2.00	1.00	2.00	0.10	9.09
ML Engineer: Adjust number of generations	24.00	7.10	170.40	45.00	6.80	306.00	0.30	4.23
ML Engineer: Deploy Oral Lesions classifier	39.00	1.70	66.30	44.00	1.70	74.80	0.00	0.00
ML Engineer: Evaluate loss curve	64.00	8.60	550.40	97.00	8.60	834.20	0.00	0.00
ML Engineer: Evaluate testing report	40.00	9.00	360.00	52.00	8.90	462.80	0.10	1.11
ML Engineer: Select the best classifier VS ML Engineer: Evaluate validation report	45.00	13.40	603.00	68.00	6.90	469.20	6.50	48.51
ML Engineer: Set development parameter	65.00	24.10	1566.50	47.00	17.20	808.40	6.90	28.63
ML Engineer: Set segregation parameter	65.00	19.00	1235.00	47.00	18.80	883.60	0.20	1.05
Provide Bounding Box and Class Label	5255.00	31.20	163956.00	5160.00	31.20	160992.00	0.00	0.00
Register hospital	1.00	12.40	12.40	2.00	11.50	23.00	0.90	7.26
Start oral lesion detection application	1.00	1.80	1.80	2.00	1.10	2.20	0.70	38.89
TOT		**206.20**	**172712.2**		**177.30**	**167376.9**	28.90	14.02
IMPROVEMENT % Total	**3.09**							
IMPROVEMENT Total Duration	**5335.30**							

After completing the TO-BE model and updating the task-level costs accordingly, we compared its costs to those of the AS-IS model.

In the table above, we analysed each task of both models by comparing the costs (Avg Duration) and the tokens processed (Count). This made it simple to calculate the difference in performance between the AS-IS model and the improved TO-BE model without considering the number of tokens processed. The last two columns show this performance difference.

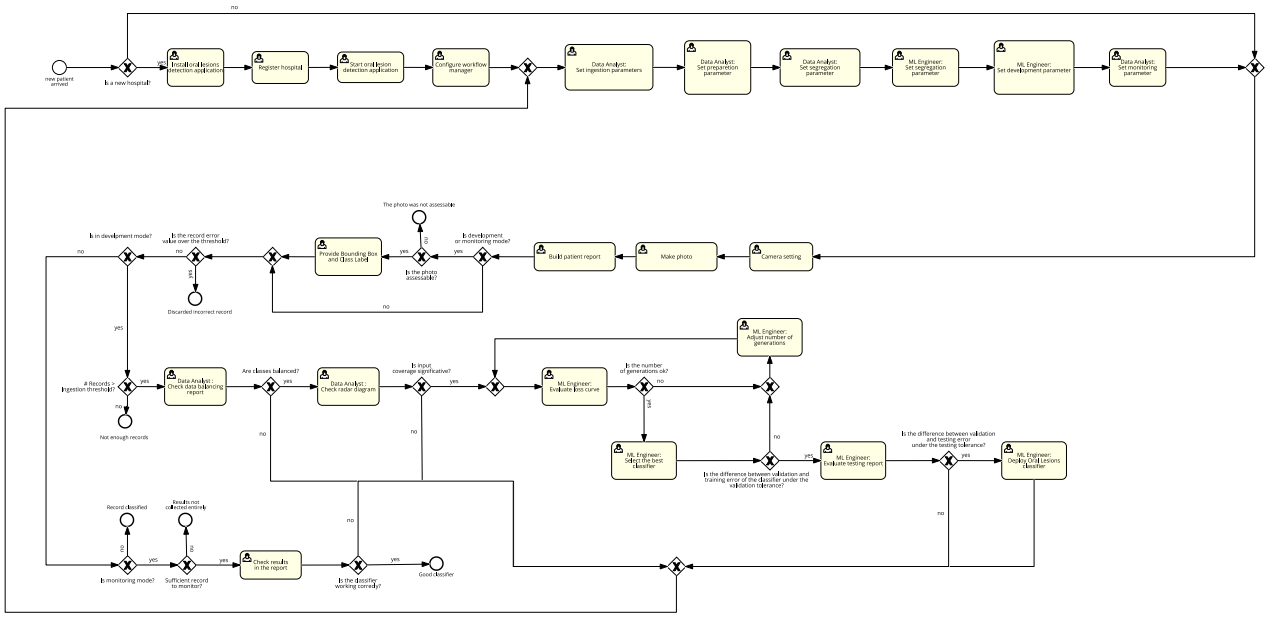
The final two columns clearly illustrate both the overall difference and the percentage difference between the two models. The percentage difference immediately highlights the impact of the improvements made by the TO-BE model on specific tasks such as "Check data balancing report" at the task level and task groups such as "Select the best classifier" and "Set development parameter" at the service level.

When making a general comparison between the two models, it is apparent that the TO-BE model had a percentage difference of 14 or 3.09 when considering the count value, in its favour. This leads us to conclude that the overall improvements made were satisfactory.

PROCESS MINING

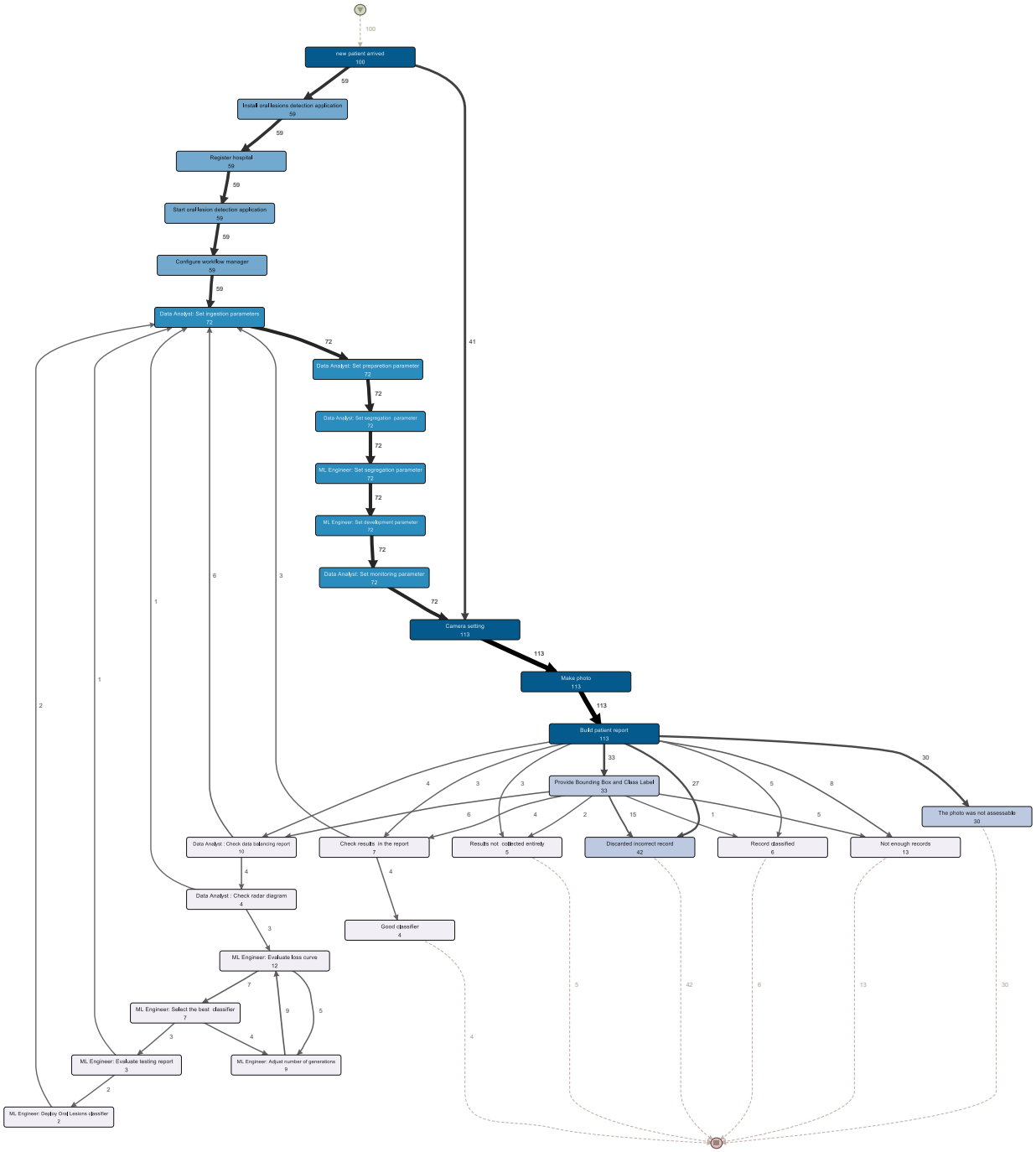
Normative model (Giuseppe)

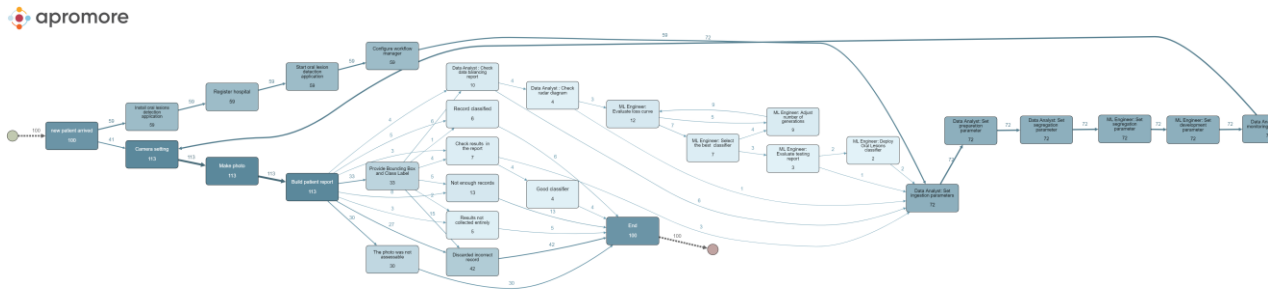
The AS-IS model is used as the normative model:



The normative model was simulated using BIMP with the following parameters: 1 euro cost and 1 sec duration to each task, 10 resources per lane, 50% to each gateway, and 100 input tokens.

Comparison between transition map generated with DISCO and transition map and with APROMORE (Giuseppe)





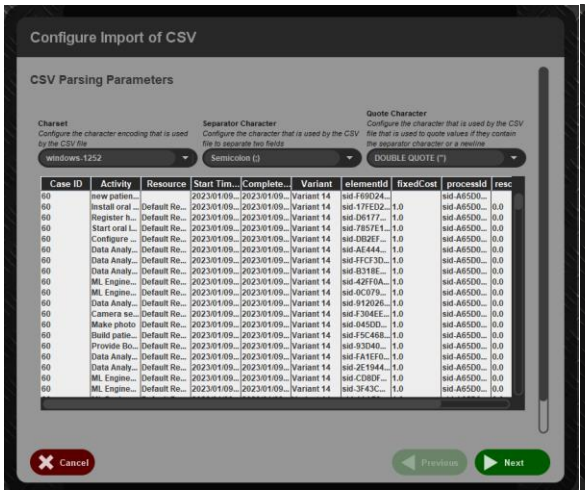
In both transition maps the highest frequencies are among the activities "camera setting", "make photo" and "build patient report", because these activities are present in all cases and in some of them these activities are repeated, (in the cases where there is the need for the annotation of additional cases and so it returns to the configuration tasks).

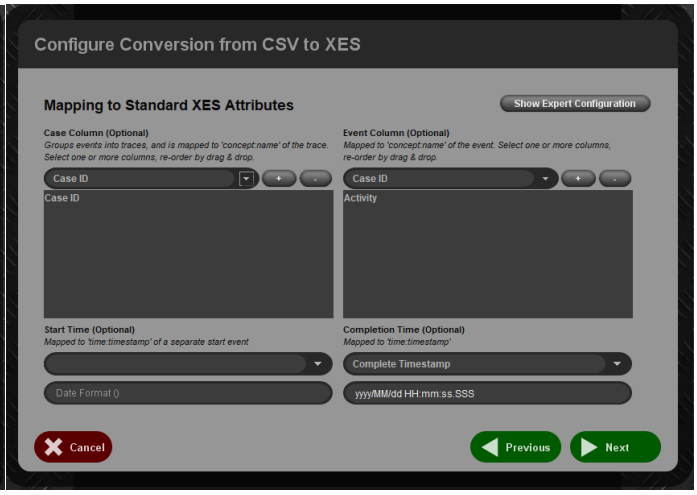
The two transition maps are the same, being generated by the same logs.

Mining on ProM Steps (Matteo)

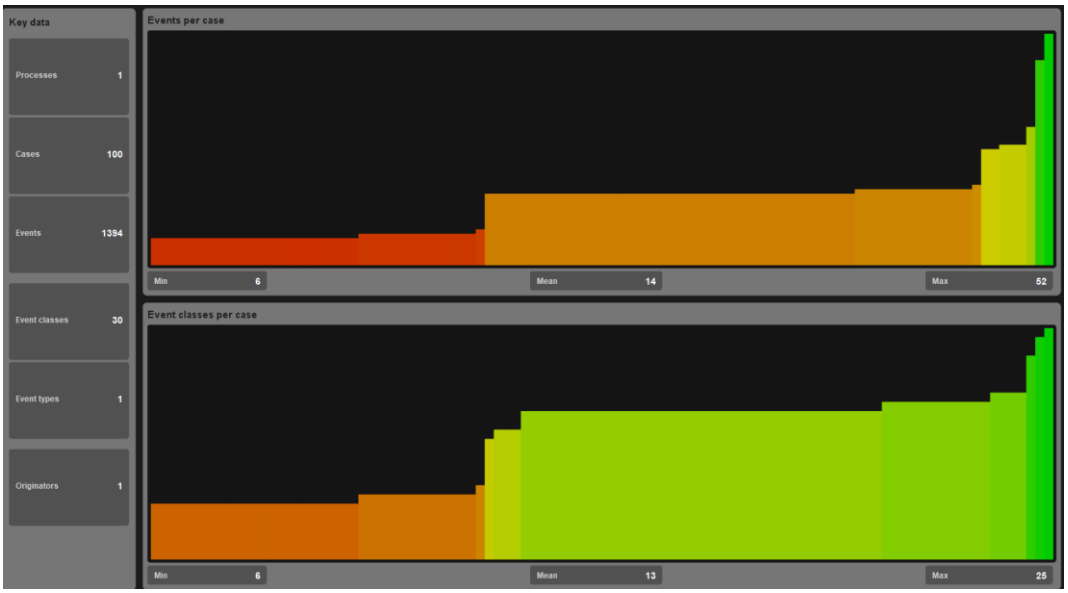
We generated logs from the AS-IS model using BIMP simulator, we inserted them into DISCO and exported it in csv adding endpoints.

We imported the log file in ProM and converted it into XES as shown in the following images.

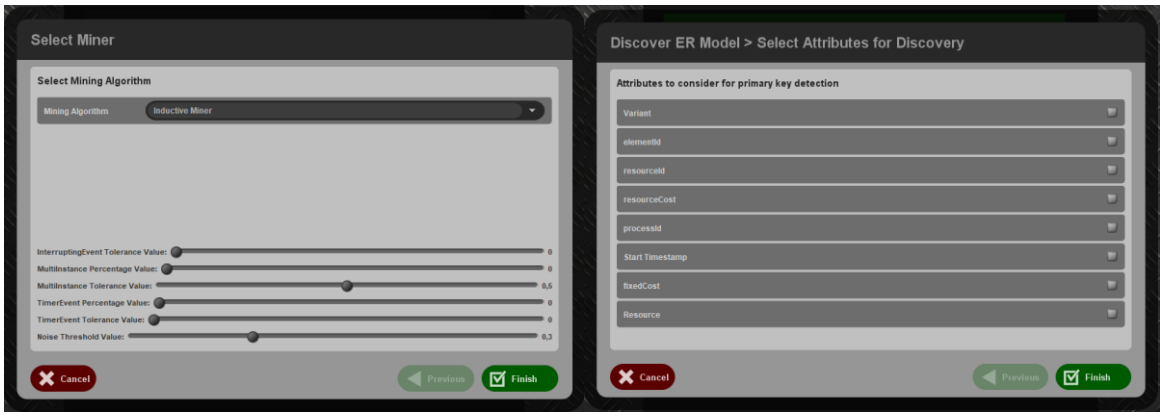




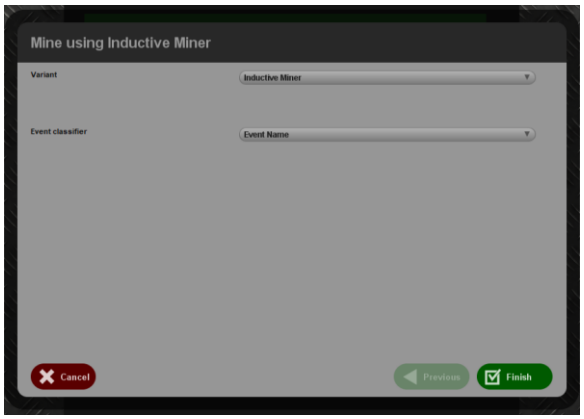
We mapped the Case Column with CaseID, the Event Column with the Activity and we set as competition time the Complete Timestamp.



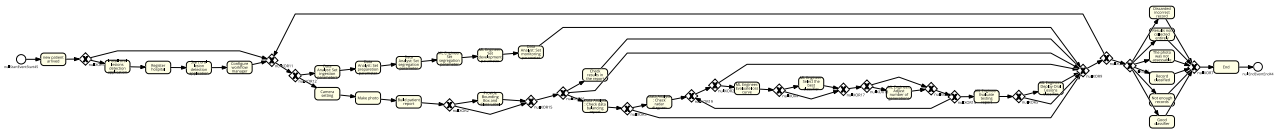
As we can see form the result, the number of cases, events and event classes correspond to the same values that we have in DISCO, so the conversion from csv and xes has been correctly completed.

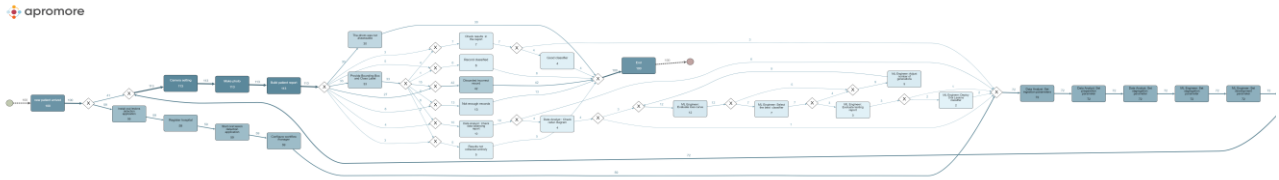


We selected the Inductive Miner and deselected all the possible primary key attributes.



Comparison between Mined BPMN from PROM and Mined BPMN from APROMORE (Giuseppe)



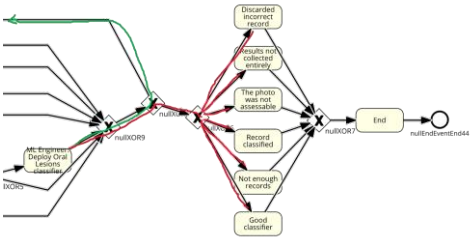


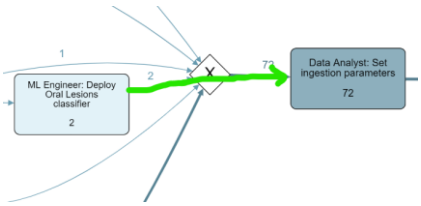
Compared to the normative model, having entered the logs in Disco and having added the endpoints, we have a single task called "End" connected to the single end event that groups all the terminations of the process.

The several normative model end events are now tasks linked to "End".

Differences between the two models (Giuseppe):

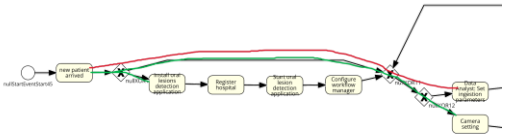
- The main difference is in how all the “return to configuration” cases is handled: in apromore, in all those cases there is a direct path to the configuration tasks (which start with the task “set ingestion parameters”), as in the normative model, while the ProM model allows several possible extra-behaviours, that are not allowed in the normative model.

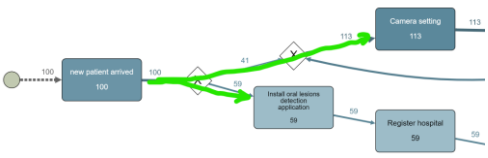




- In the ProM model, we might have cases who have a direct path to the end, they can continue with activities from the configuration phase.
  For example a case where we have the "Build patient report" activity followed by "set ingestion parameters" or "camera setting" would not violate the model extracted from ProM, but would violate the normative model or the one extracted from apromore.

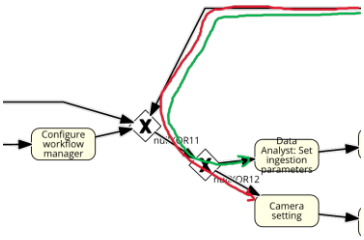
- In apromore after "new patient arrived" the only two possible paths are "Install oral lesion application" and "camera setting", while in ProM it would also be possible a case that after "new patient arrived" goes to "set ingestion parameter", behavior that is not allowed in the normative model.





- There could be cases in which a sequence of activities that should continue with configuration activities, such as "set ingestion parameters", continues instead with the "camera setting" activity,

without violating the ProM model but violating the normative model and the one extracted from apromore.



Therefore from this first visual comparison we expect the model extracted from ProM to be less precise than that of apromore, given that it would allow more extra-behaviours.

We calculated using the ProM plugins Fitness, Precision, Generalization and Simplicity in order to compare the two models mined with ProM and apromore.

We calculated the simplicity as the sum of ***# of gateways + # of sequence flows + # of activities***.

Conformance check on original log	Fitness	Precision	Generalization	Simplicity
BPMN ProM	1,0	0,73244	0,98071	19 + 30 + 29 = 78
BPMN APROMORE	1,0	0,86282	0,98071	20 + 30 + 29 = 79

As seen before in the comparison between the two models, it is confirmed that the model extracted with ProM is less precise. Anyway, we expected a difference in the Generalization between the two models.

Since Fitness and generalization are the same, in our opinion the best model is the **apromore** one, because although it's more complex than the ProM one (slightly more complex: 79 vs 78), in our opinion the clearly higher precision can justify the higher cost given by the slightly higher complexity.

LOG Violation (Mattia Di Donato)

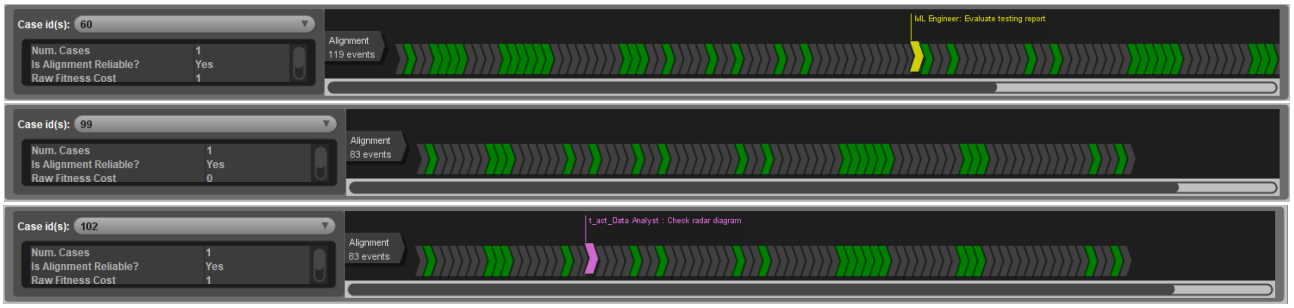
We incorporated three realistic violations into our company's BPMN model, which affected the Configuration, Annotation, Ingestion, Preparation, Segregation, and Development systems.

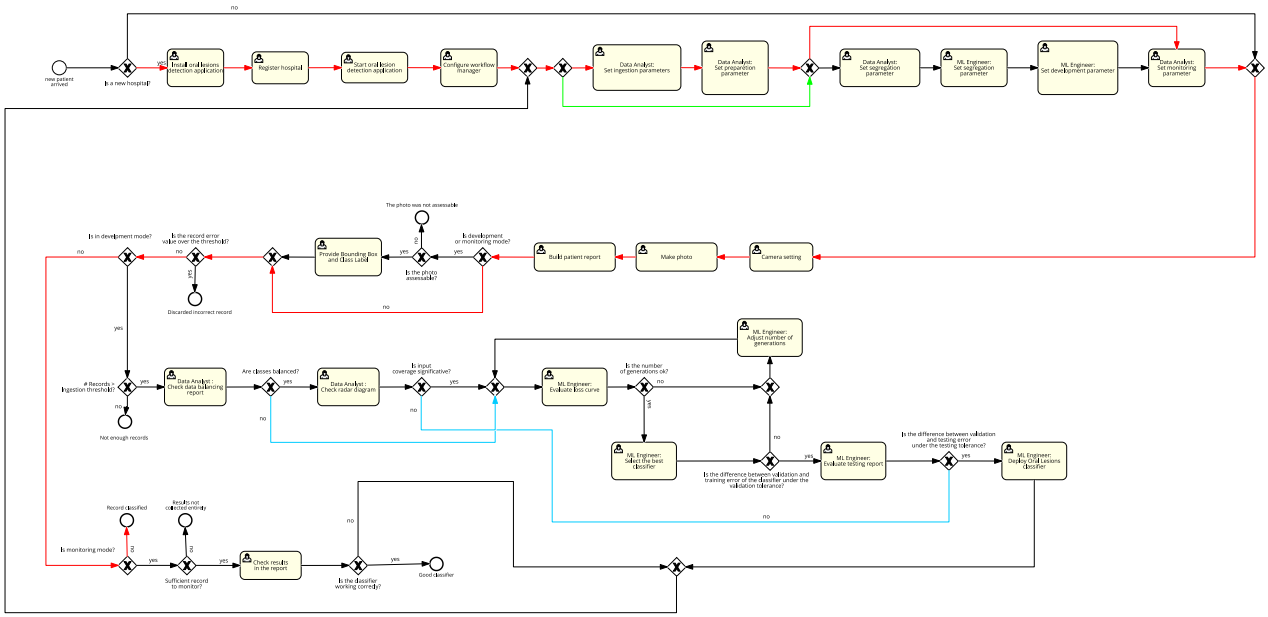
Specifically, the following features were introduced:

1. Shared use of a single classifier for similar hospitals (red): We proposed using the same classifier for similar hospitals, which would eliminate the need for the Development and Segregation System, as data collection and training would no longer be necessary. Additionally, part of the configuration process related to the definition of Development and Segregation parameters could be skipped. Following this violation, we update the logs 100,69 skipping the configuration tasks and the log 73 skipping configuration and provide bounding box tasks.

2. Automatic configuration of the Ingestion and Preparation systems (green): We proposed automating the configuration of the Ingestion and Preparation systems. This violation includes switching the operating mode from execution to development automatically, and using default values for all other parameters, making them the same for all.
Following this violation, we update the logs 101, 62 and 25.

3. Additional records from other hospital (blue): If similar hospitals can share records, the production process could work without returning to the configuration phase if data balancing, coverage of the radar diagram, or when the difference between validation and testing error is over the testing tolerance. In this case, the system would automatically request new data from similar hospitals to improve its own dataset.
Following this violation, we update the log 102 to emulate unbalanced class event, log 99 to emulate error in input coverage and log 60 to emulate error under testing report.




new patient arrived
Is a new hospital?
yes
no
Install oral lesions detection application
Register hospital
Start oral lesion detection application
Configure workflow manager
Data Analyst: Set ingestion parameters
Data Analyst: Set preparation parameter
Data Analyst: Set segregation parameter
ML Engineer: Set segregation parameter
ML Engineer: Set development parameter
Data Analyst: Set monitoring parameter
Is in development mode?
Is the record error value over the threshold?
Discarded incorrect record
Provide Bounding Box and Class Label
The photo was not assessable
Is the photo assessable?
Is development or monitoring mode?
Build patient report
Make photo
Camera setting
Records > Ingestion threshold?
Not enough records
Data Analyst : Check data balancing report
Are classes balanced?
Data Analyst : Check radar diagram
Is input coverage significative?
ML Engineer: Evaluate loss curve
Is the number of generations ok?
ML Engineer: Adjust number of generations
ML Engineer: Select the best classifier
Is the difference between validation and training error of the classifier under the validation tolerance?
ML Engineer: Evaluate telling report
Is the difference between validation and training error of the classifier under the testing tolerance?
ML Engineer: Deploy Oral Lesions classifier
Is monitoring mode?
Record classified
Results not collected entirely
Sufficient record to monitor?
Check results in the report
Is the classifier working correctly?
Good classifier

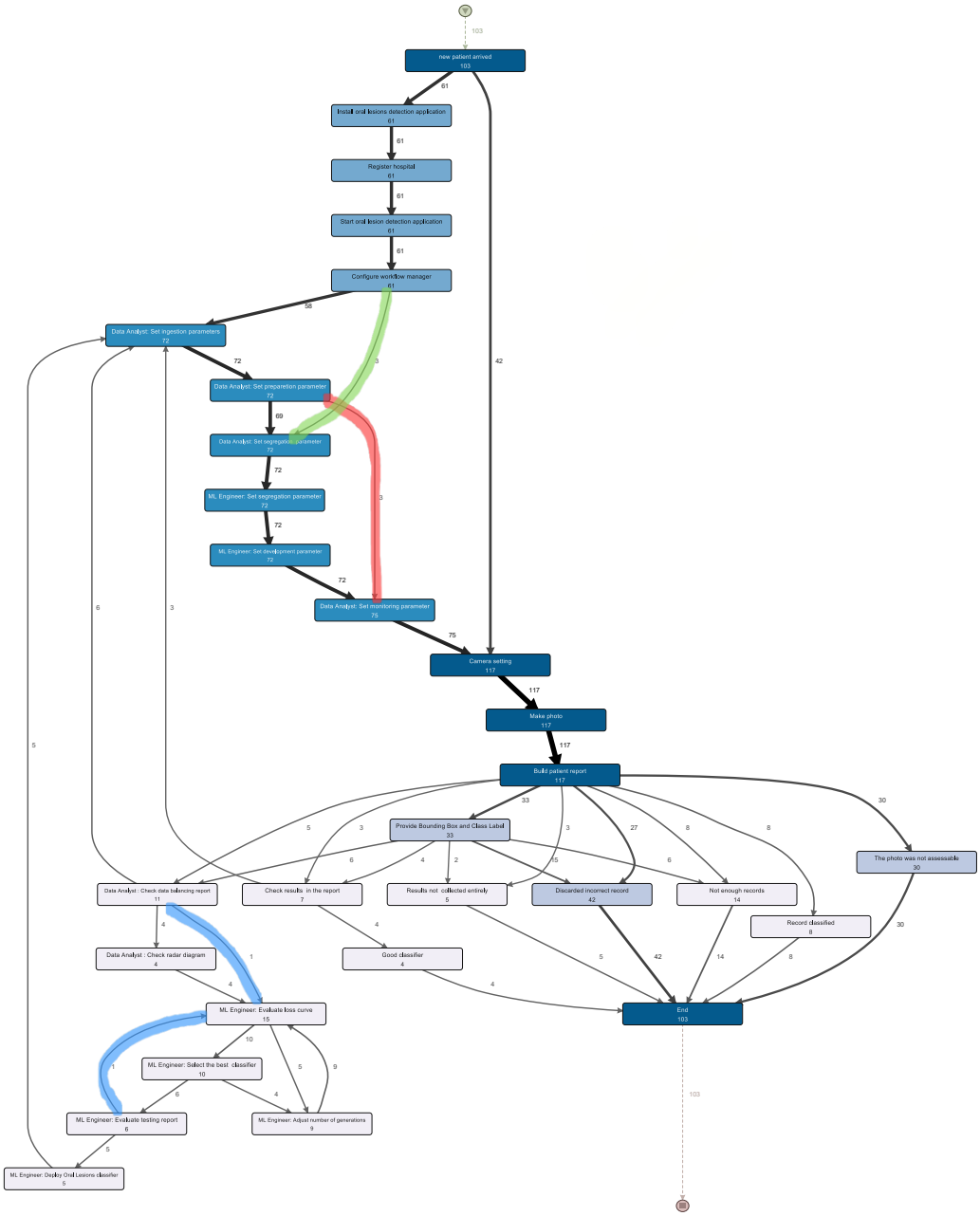
Transition Map of violated logs(Salvatore)

In the following image, the logs that violate the normative model have been highlighted.

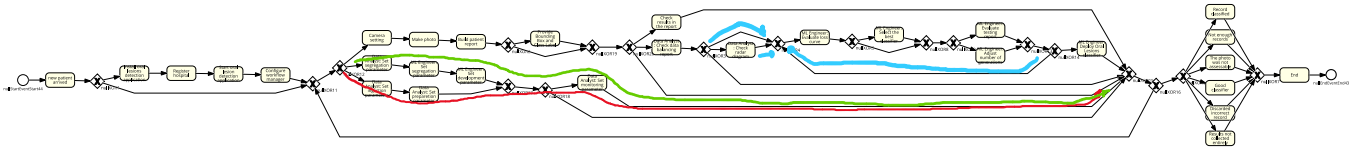
The green edge corresponds to all 3 logs (1 added and 2 modified) and therefore three tokens, which for the reasons already described above, jump directly from the "configure workflow manager" task to "set segragation parameter".

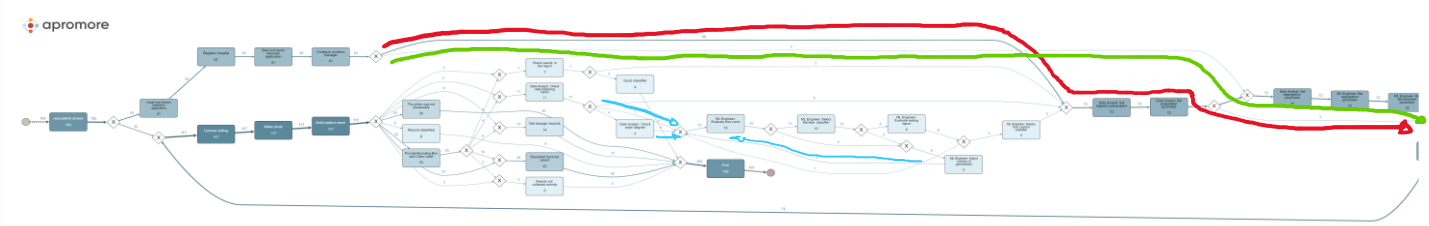
The red edge corresponds to all 3 logs (1 added and 2 modified) and therefore three tokens, which for the reasons already described above, jump from "set preparation parameter" to "set monitoring parameter".

As for the edges connected in blue, they should be three but in this case, the violated log that goes from "check radar diagram" to "evaluate loss curve" still follows a path that is accepted in the normative model and therefore is not considered as a violation.



Comparison between BPMN mined from modified logs using ProM and BPMN mined from modified logs using Apromore (Matteo)





As we expected the fitness measure of the modified model is 1.0 respect to the violated log instead in the original models, it is less than 1 because they can't perfectly replay the event log.

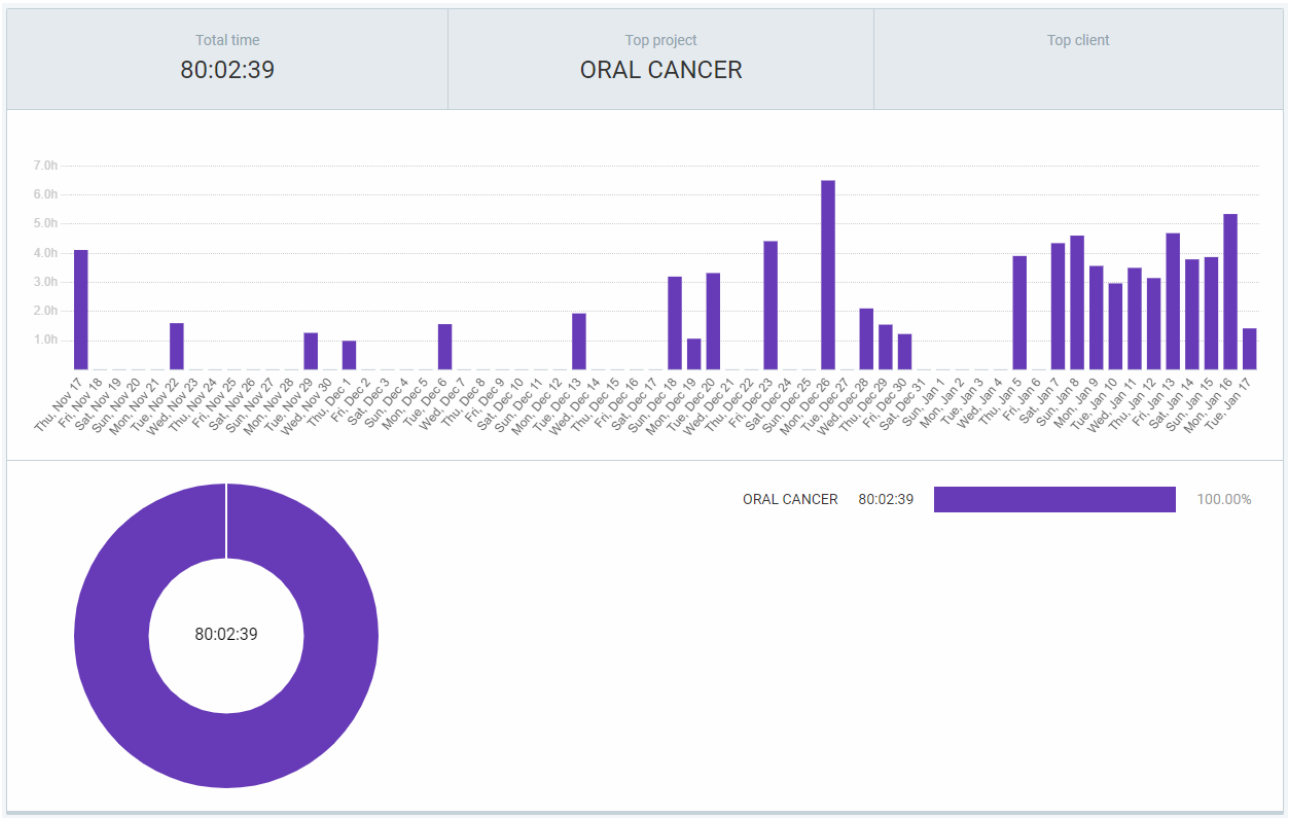
Generally better-quality indicators are obtained with Apromore models, indeed if we focus our attention on the modified BPMN we can see that the ProM we have the possibility to skip the setting of the monitoring parameter, a behavior which was not present in the modified log, while in the Apromore model is not possible to pass through the configuration phase without setting the monitoring parameters. This difference makes the Apromore BPMN more precise respect to the ProM one. We obtain the same generalization value even though we thought to obtain an higher value on the ProM ones because they are less precise. The simplicity is the sum of number of gateways, number of sequence flow and number of activities.

Conformance check on modified log	Fitness	Precision	Generalization	Simplicity
BPMN ProM from original log	0,99197	0,72417	0,98677	19 + 30 + 29 = 78
BPMN APROMORE from original log	0,99173	0,85597	0,98859	20 + 30 + 29 = 79
BPMN ProM from modified log	1,0	0,63533	0,98427	19 + 33 + 30 = 82
BPMN APROMORE from modified log	1,0	0,79951	0,98427	22 + 33 + 30 = 85

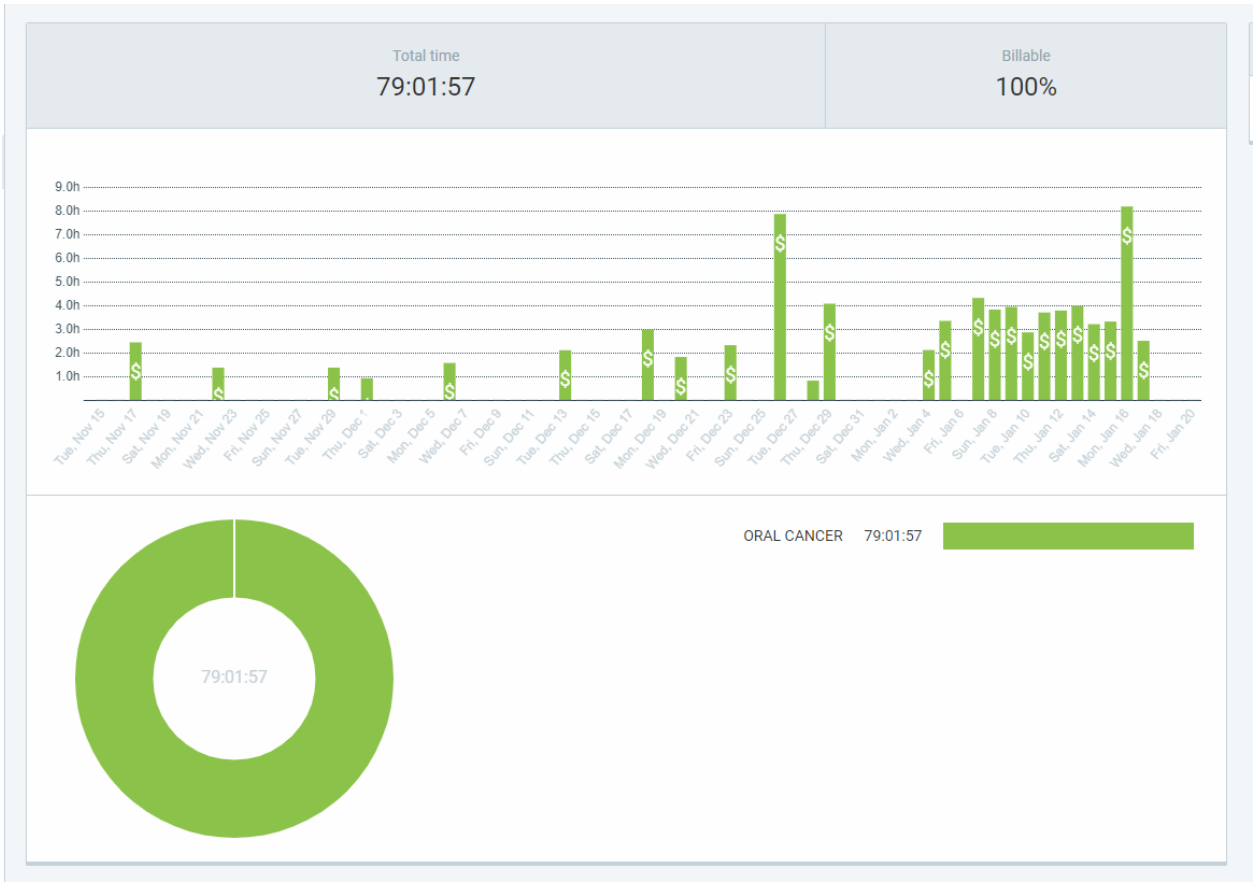
In conclusion for our logs it seems that the Apromore solution is more precise but more complex to understand while the ProM solution should be able to generalize better and be more simple to understand.

Clockify Report

Salvatore Arancio Febbo

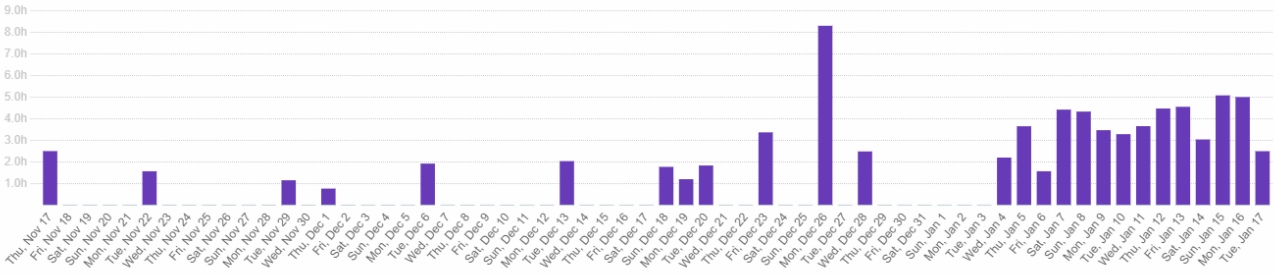


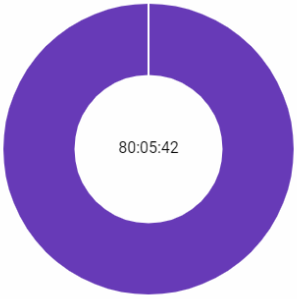
Mattia Di Donato



Matteo Giorgi

Total time	Top project	Top client
80:05:42	ORAL CANCER	





Giuseppe Martino

Total time	Top project	Top client
80:09:50	ORAL CANCER