CIGI QUALITA MOSIM 2023

Surgical Duration Prediction by Developing a Hybrid Model (Using Machine Learning Techniques)

SHAHAB AMROLLAHIBIOUKI[1], PROF. YVAN BEAUREGARD[1]

[1] Mechanical Engineering Département, École de Technologie Supérieure
Montréal, H3C1K3, Canada
Shahab.Amrollahibiouki.1@ens.etsmtl.ca, Yvan.Beauregard@etsmtl.ca

***Résumé* – La majorité des revenus et des coûts d'un hôpital sont liés à ses salles d'opération; il est donc crucial pour l'administration hospitalière de planifier et de programmer ce service afin d'optimiser son efficacité. Ils ont besoin d'informations fiables sur la durée de chaque intervention chirurgicale pour gérer la salle d'opération de manière précise et efficace. La plupart des estimations de durée de la chirurgie sont empiriques et basées sur l'expérience du chirurgien, ce qui est souvent imprécis. Par conséquent, nous pouvons prédire la durée de la chirurgie de manière plus précise en appliquant des méthodes scientifiques telles que des algorithmes d'apprentissage automatique, et sur cette base, nous pouvons planifier les salles d'opération de manière plus précise. Ainsi, cette étude évalue une méthode de prédiction de durée de chirurgie basée sur l'apprentissage automatique. Pour prévoir la durée de la chirurgie au Centre universitaire de santé McGill à Montréal, au Canada, nous utilisons et évaluons quatre techniques d'apprentissage automatique (régression linéaire, machines à vecteurs de support, arbres de régression et arbres bagués). Les algorithmes seront entraînés à l'aide de données historiques du système de dossier médical électronique de ce centre de santé. L'erreur quadratique moyenne (RMSE) de la durée chirurgicale prédite et le temps de calcul des algorithmes sont utilisés pour comparer les performances des méthodes. Pour améliorer la précision des prévisions, nous proposons un modèle hybride en plus des modèles classiques de prédiction. Enfin, nous comparons l'efficacité des quatre algorithmes d'apprentissage automatique utilisés pour estimer la durée de la chirurgie à notre modèle hybride. Sur la base des résultats, nous choisissons la méthode qui offre la meilleure performance tout en maintenant un compromis raisonnable avec l'une des plus courtes RMSE et des coûts de calcul raisonnables. Ensuite, nous déterminons quel modèle permettra d'estimer la durée de la chirurgie de la meilleure manière possible.**

***Abstract* -The majority of a hospital's revenues and costs are tied to its operating rooms; hence it is crucial for hospital administration to plan and schedule this department to optimize its effectiveness. They require reliable information about how long each surgery will take to manage the operating room precisely and effectively. Most surgical duration estimates are empirical and based on the surgeon's experience, which is frequently inaccurate. Therefore, we can predict the duration of surgery more correctly by applying scientific methods like machine learning algorithms, and based on this, we can plan operating rooms more precisely. As a result, This study evaluates a machine learning-based surgery time prediction method. To forecast the duration of surgery at The McGill University Health Centre in Montreal, Canada, we use and assess four machine-learning techniques (Linear Regression, Support Vector Machines, Regression Trees, and Bagged Trees). The algorithms will be trained using historical data from this health center's Electronic Health Record system. The Root Mean Square Error (RMSE) of the predicted surgical duration and the computation time of the algorithms are used to compare the performance of the methods. To improve forecasting accuracy, we provide a hybrid model in addition to typical models for prediction. Finally, we compare the effectiveness of the four ML algorithms used to estimate surgery duration to our hybrid model.Based on the results, we choose the method that provides the highest performance while maintaining a reasonable compromise with one of the shortest RMSEs and computational costs. Then we determine which model will estimate the duration of the surgery the best.**

Mots-clés – Durée de la chirurgie, Machine learning, Modèle hybride
***Keywords* – Surgery Duration, Machine Learnin, Hybrid Model**

1 INTRODUCTION

In healthcare, the operating room (OR) plays a crucial role as it is the location where surgeries are performed to diagnose and treat a wide range of medical conditions. The OR is a complex environment that requires coordination and communication among multiple healthcare professionals, including surgeons, nurses, and anesthesiologists, to ensure that surgeries are performed safely and efficiently. One of the key aspects of managing the OR is surgery scheduling [Fairley et al., 2019].

Surgery scheduling involves allocating OR time to different surgeries and ensuring that the necessary personnel and equipment are available for each procedure. Accurate scheduling is important to ensure that surgeries are performed in a timely manner and that patients receive the care they need as quickly as possible [Cardoen et al., 2009]. Additionally, accurate predictions of surgery duration are also important in the OR management [Putra Twinanda et al., 2019]. Knowing how long a surgery is expected to take allows for better planning and scheduling of surgeries, as well as more efficient use of OR resources. Accurate surgery duration predictions can also help to reduce the time patients spend in the OR, which can lead to shorter recovery times and improved outcomes [Raymond et al., 2019].The efficient use of operating room resources is crucial for ensuring that surgeries are completed on time and without delays. However, this can be difficult to achieve if the operating room schedule is not properly managed. If the operating room schedule allocates less time for a surgery than the actual duration, the surgery will not begin on time. Conversely, if the planned surgery length is longer than the actual operation duration, the operating room will have to remain empty, which can result in a waste of resources. To avoid these issues, it is important for operating room administrators to accurately estimate the length of surgeries and anesthetic emergence, and to schedule operating rooms accordingly. This can be done by using prior experience or by basing the estimation on the type of surgery and other relevant factors. By managing the operating room schedule in this way, the resource consumption can be reduced and surgeries can be completed in a more efficient manner [Dexter et al., 199] [Amrollahibiouki et al., 2021].

However, this prediction strategy is a relatively sloppy and unscientific assessment of a wide range, which frequently results in significant inaccuracies in the duration of the surgery and the emergence of anesthesia and wastes resources in the operating room. In a few earlier studies, researchers chose a few variables and applied a straightforward multiple regression method to forecast how long specific processes would take, but this approach has some drawbacks as well.

Numerous machine learning techniques have recently been created and successfully used for a variety of prediction problems. [Master et al., 2017] shown that tree-based machine learning algorithms trained with surgeons' estimates of surgery length may produce noticeably better predictions of the duration of pediatric procedures (with a series of surgeries taking an average of 44 minutes, and 60% of average accuracy.) than averaging historic duration times or surgeon' estimates alone.

Additionally, [Davies et al., 2004] found that the K-Nearest Neighbors algorithm performed poorly in comparison to the other three models, with a higher error rate in predicting the occupation time of the operating room. This highlights the importance of selecting the appropriate machine-learning algorithm for a specific task.

Furthermore, the results of the study suggest that incorporating a wide range of data from the patient's medical history can significantly improve the accuracy of predictions. This highlights the importance of considering multiple factors when making predictions, rather than relying solely on a few specific factors.

Overall, the study by Davies provides valuable insights into the performance of different machine learning algorithms in predicting the occupation time of the operating room. It emphasizes the importance of selecting the appropriate algorithm and incorporating a wide range of data when making predictions.

Additionally, [Fairley et al., 2019] found that incorporating heuristics algorithms, such as deterministic optimization, in combination with the Regression Tree algorithm led to an improvement in the accuracy of predictions. This suggests that combining different types of algorithms can lead to better performance in certain tasks.

Furthermore, the study highlights the importance of considering multiple factors when making predictions. The use of ten input factors, including age, service, patient class, and gender, led to a more accurate prediction of surgery durations. This supports the findings of [Davies et al., 2004] that incorporating a wide range of data can significantly improve the accuracy of predictions.Overall, the study by Fairley et al. Demonstrates the effectiveness of using machine learning estimators with heuristics algorithms for predicting the occupation of a hospital's Recovery Unit. It also highlights the importance of considering multiple factors and combining different types of algorithms for improved performance.

[asebrook et al., 2017] highlights the potential of using a Linear Regression technique for forecasting the surgery duration of procedures in different university hospitals. By using only a few factors such as the patient's age, the kind of anesthetic, and pre-surgical length predictions, the study was able to achieve modest prediction errors of around 39 minutes in a group of operations with a mean Total Procedure Time of 150 minutes.

Additionally, the study by [Hosseini et al., 2015] used data mining techniques to extract variables from the hospital's clinical histories to estimate the surgical time for 15 specialties. The study employed classical least square linear regression (LIN) and stepwise regression (STEP) to analyze 63,000 procedures over the course of 39 months. The input factors used in the study included surgery type, procedure type, physical state, patient age, surgical scope, and specialty.

Predicting the duration of surgery is an important aspect of healthcare as it allows physicians and hospitals to plan and schedule surgeries efficiently. However, predicting the duration of surgery can be a challenging task as it is affected by various factors such as the patient's health condition, type of surgery, and surgeon's experience. In recent years, machine-learning algorithms have been used to predict the surgery duration with increasing accuracy. However, many of these algorithms rely on a single machine-learning technique which can limit their accuracy and have limitations, such as poor generalization, lack of interpretability, and high computational complexity. Therefore, to overcome these limitations, we develop a hybrid machine-learning algorithm that combines multiple machine-learning techniques to predict the surgery duration. By using a combination of supervised and unsupervised learning techniques, we aim to increase the accuracy of the predictions. This accurate surgery duration will not only help in better scheduling and planning of surgeries but also help in reducing overall healthcare costs.

The hybrid machine learning algorithm is trained by data that will be obtained from records of planned and actual surgery times performed by The McGill University Health Centre in Montreal, Canada. For the surgical time (duration) forecasting model, several tests will be run to improve the design of the proposed hybrid model that works effectively based on the Mean Square Error and the training and validation phases.

2 METHODS

The method described in this study aims to develop a hybrid model for predicting the duration of surgeries. The model combines the strengths of supervised and unsupervised learning algorithms to achieve improved prediction accuracy. Specifically, we will be using the random forest algorithm for supervised learning, which utilizes a combination of decision trees to predict the surgery duration based on selected features.

Additionally, we will be using the k-means algorithm for unsupervised learning. The k-means algorithm is a popular clustering technique, which is used to group similar data points together based on their characteristics. In this study, k-means will be used to group patients based on their demographic information, medical history, and other relevant features. By grouping similar patients together, we can identify patterns and trends that may not be apparent when looking at individual patients.

To calculate the distance between patients, we will be using Euclidean distance. Euclidean distance is a commonly used distance metric in k-means and other clustering algorithms. It is calculated as the square root of the sum of the squares of the differences between the coordinates of two points. This distance metric is chosen because it is simple to compute, and it is sensitive to the difference in all dimensions.

We chose k-means as a clustering algorithm because it is a robust and widely used algorithm in machine learning. K-means is simple to implement and easy to interpret, which makes it suitable for our study. By using k-means, we can identify patterns and trends that may not be apparent when looking at individual patients, which will help in predicting the duration of surgeries

2.1. Data collection:

The first step in developing a hybrid machine-learning algorithm to predict the surgery duration is to collect a large dataset of patients who have undergone surgery at The McGill University Health Centre (MUHC). This dataset should include information such as patient demographics, pre-surgery health conditions, type of surgery, and actual surgery duration. The data can be collected from the electronic medical records (EMR) of the patients. To ensure consistency in operating procedures and relevant data, the data collection period for this study will be defined based on the following general guidelines:

- Data should be recent enough to be relevant to current surgical practices and patient characteristics. This ensures that the model developed in this study will be applicable to current patients and surgeries.
- Data should be available for a sufficient number of patients to ensure an adequate sample size for the study. This ensures that the model developed in this study will be generalizable to the population.
- Data should be diverse enough to include a variety of surgical procedures and patient characteristics. This ensures that the model developed in this study will be able to handle a wide range of surgeries and patient types.
- Data should be reliable and well-documented to ensure accurate analysis and conclusions. This ensures that the model developed in this study will be based on accurate data and that any conclusions drawn from the study will be valid.

2.2. Feature selection:

In order to improve the accuracy of our prediction models for surgery duration, it is important to select the most relevant variables from the dataset. Incorporating knowledge experts through interviews can provide valuable insights into which variables are likely to have the greatest impact on surgery duration.

In our study, we conducted interviews with practicing surgeons, anesthesiologists, and other medical professionals who have extensive experience in the surgical field. These experts provided valuable input on the variables that are most likely to impact surgery duration, such as patient age, BMI, medical history, type of surgery, and anesthesia type.

Based on their insights, we identified several variables in the dataset that are most relevant for predicting surgery duration, including patient age, surgeon age, anesthesia type, the number of personnel present in the operating room, BMI, surgery hours, days in the hospital before surgery, surgeon specialty, admission department, and hospitalization type. These variables were selected based on their potential impact on surgery duration and their availability in the electronic medical records. The variables in this data set that we require to focus on are listed in Table 1 :

Table 1. Variables in dataset

Variable	Definition
Patient Age	Age of patient (Numeric)
Surgeon Age	Age of Surgeon (Numeric)
K Surgeon	K Surgery
K Anesthésia	K Anesthesia
No. Personnel Surgery	The number of personnel present in the operating room
BMI	Body mass index (BMI)
Surgery Hours	The time the surgery is performed
Anesthesia System	Anesthesia method
Days in Hospital	The duration of the patient's hospital stay before surgery
Specialty	Surgeon specialty
Admission	The department where the patient is admitted
Hospitalization type	Hospitalization type

Once the relevant variables have been identified, the next step is to use stepwise regression to select the most important features for training the prediction algorithm. By using stepwise regression, we can identify the most important features that have the most impact on the surgery duration, and use only these features to train the algorithm for improved prediction accuracy

2.3. Data preprocessing:

The selected features will then be preprocessed to ensure that they are in a format that can be used by the machine-learning algorithm. This can include normalizing the data, removing outliers, and handling missing values.

2.4. Supervised learning:

Training a supervised learning algorithm such as a random forest on the preprocessed data using the selected features. Use a portion of the dataset for training and the remaining data for testing. The random forest algorithm will be used to predict the surgery duration based on the selected features.

2.1.1 Random forest algorithm

The predicted surgery duration using the random forest algorithm will depend on the specific dataset and features that are used for training and testing. However, in general, the random forest algorithm is known for its ability to accurately predict outcomes based on multiples inputs variables (Huang et al., 2020). It also has the advantage of being able to handle large datasets and non-linear relationships between variables. Therefore, it is likely that the predicted surgery duration will

be accurate and consistent with the actual duration of the surgeries in the dataset.
The mathematical formula for predicting the surgery duration using a random forest algorithm is based on the decision trees that make up the forest. Each decision tree in the forest is trained on a random subset of the data and features, and the final predicted surgery duration is determined by averaging the predictions of all the trees in the forest.
The decision tree algorithm is a recursive partitioning method that creates a tree-like structure of decisions based on the input features. Each node in the tree represents a decision based on the value of one of the input features, and the branches represent the possible outcomes of that decision. The leaves of the tree represent the final predicted outcome.
In the case of predicting surgery duration, the input features would be the selected variables that are believed to have an impact on the duration of the surgery. For example, these could include the type of surgery, the patient's age and health condition, and the experience of the surgeon.
The decision tree algorithm would start at the root node, which represents the overall dataset. It would then create splits in the data based on the values of the input features, with each split leading to a new decision node. This process would continue until the tree reaches a final leaf node, which represents the predicted surgery duration.
To calculate the predicted surgery duration using a random forest, the algorithm would create multiple decision trees, each one trained on a different subset of the data and features. The final predicted surgery duration would be determined by averaging the predictions of all the trees in the forest.

2.5. Unsupervised learning :

Use an unsupervised learning algorithm such as k-means clustering to identify patterns in the data. The k-means algorithm will be used to group the patients based on their demographics, pre-surgery health conditions, and type of surgery.

2.1.2 k-means algorithm

The k-means algorithm is an unsupervised learning method that is used to group similar data points together in clusters [Chowdhury et al., 2021]. It works by first initializing a specified number of "centroids," or cluster centers, in the data. The algorithm then assigns each data point to the cluster center that is closest to it, based on a distance metric. The cluster centers are then recalculated as the mean of all the points in the cluster, and the process is repeated until the cluster assignments no longer change.
In the context of identifying patterns in patient data, the k-means algorithm could be used to group patients based on their demographics, pre-surgery health conditions, and type o of surgery. These variables could be chosen as the input features for the algorithm, and the k-means algorithm would group patients with similar values for these features together in the same cluster.
For example, the algorithm could group patients based on their age, gender, and pre-surgery health conditions, such as diabetes or hypertension. This would allow for the identification of patterns in the data such as which demographics or health conditions are more commonly associated with specific types of surgery.
Another example, the k-means algorithm could be used to group patients based on the type of surgery they underwent and the length of the surgery. This would allow the for identification of patterns in the data such as which surgeries tend to be shorter or longer than others.

2.6. Combining the predictions :

We combine the predictions made by the supervised and unsupervised learning algorithms to make a final prediction for the surgery duration. The supervised portion of the algorithm, specifically the random forest algorithm, will be used to predict the surgery duration based on selected features. The unsupervised portion of the algorithm, specifically the k-means algorithm, will be used for clustering patients based on their characteristics. The final prediction for the surgery duration will be made by combining the predictions from both the supervised and unsupervised portions of the algorithm .

2.7. Model evaluation:

Compare the predicted surgery duration to the actual surgery duration to evaluate the performance of the model. Metrics such as mean absolute error and mean the squared error will be used to evaluate the performance of the model. In addition to the mean absolute error and mean squared error, we will also use metrics such as the coefficient of determination (R-squared) and the root mean squared error to evaluate the performance of the model. These metrics will provide a more comprehensive understanding of the model's performance, including the degree of error, the proportion of variation explained by the model and the overall goodness of fit of the model.

2.8. Model deployment :

Deploy the model in a clinical setting at the MUHC to predict the surgery duration for new patients.
By collecting data from the MUHC and using a hybrid machine-learning algorithm, this project aims to increase the accuracy of the predictions for surgery duration. The use of both supervised and unsupervised learning techniques, along with the k-means clustering algorithm, can help in identifying patterns in the data that might not be captured by supervised learning alone, leading to more accurate predictions.

3 PREDICTION MODEL

Surgical duration is a crucial factor in the medical field, and accurate prediction of surgical time can help improve patient outcomes, optimize resource utilization, and reduce costs. However, predicting surgical duration is a complex task that depends on several variables, including patient characteristics, surgical procedures, surgeon experience, and more. Therefore, we propose a hybrid model that combines K-means clustering and Random Forest regression to leverage the benefits of both techniques.

3.1. k -means for Algorithm in Surgical Duration Prediction

K-means Techniques for Surgical Duration Prediction Surgical duration prediction is an important task in healthcare operations management, as it helps optimize scheduling and resource allocation in surgical units. However, the presence of heterogeneity in surgical datasets can lead to incorrect model building and predictions. To address this issue, we employ K-means clustering, an unsupervised machine-learning technique, to partition the surgical dataset into clusters based on their similarities. This approach makes the data more homogeneous within clusters and heterogeneous between clusters, thereby reducing the effects of unobserved heterogeneity. K-means clustering works by selecting an initial set of k-cluster centroids and assigning data points to the nearest centroid based on their Euclidean distance. The algorithm then computes the mean of each data point in each cluster to find its new centroid and repeats this process until each point is assigned to its nearest cluster. The primary goal of K-means clustering is to minimize the squared error between each data point and its assigned centroid. This optimization objective is represented by the following equation 1 :

$$(1)\quad f(x) = \sum_{n=1}^{k} \sum_{n=1}^{n} \left| X_i - C_j \right|^2$$

where k is the number of clusters, n is the number of cases, Cj is the number of centroids, and X is the data points from which the Euclidean distance from the centroid is calculated.

By using K-means clustering in combination with classification techniques, we can build more accurate and efficient surgical duration prediction models. Specifically, clustering the surgical dataset helps reduce the computational memory required to build models on massive amounts of data, while classification techniques can be used to train models with shorter processing times and higher accuracy.

3.2. Random Forest Algorithm in Surgical Duration Prediction

Surgical duration prediction is a critical task in healthcare operations management, as it helps optimize scheduling and resource allocation in surgical units. Various machine learning techniques have been proposed to address the complexity and heterogeneity of surgical procedures, including Random Forest algorithm.

Random Forest algorithm, proposed by Breiman and Adele Cutler, is an ensemble learning method that constructs multiple decision trees and aggregates their results to achieve high accuracy in prediction tasks. It has been widely used in various fields, including medical applications such as surgical duration prediction.

In surgical duration prediction, Random Forest algorithm is an effective method to handle the complexity and heterogeneity of surgical procedures, which can be affected by many factors such as patient condition, surgical technique, and surgical team. Random Forest algorithm can handle high-dimensional and noisy data and prevent overfitting, which makes it a suitable method for surgical duration prediction.

The Random Forest algorithm consists of the following steps:

1. Randomly select a subset of features from the dataset.
2. Construct a decision tree based on the selected subset of features.
3. Repeat steps 1 and 2 to construct multiple decision trees.
4. Aggregate the results of the decision trees to make a final prediction.

In the Random Forest algorithm, the feature subset selection and decision tree construction are performed randomly, which ensures the diversity of the decision trees and reduces the risk of overfitting. The aggregation of the results from multiple decision trees can improve prediction accuracy and stability.

Mathematically, the Random Forest algorithm can be expressed as follow :

- X: The dataset with n instances and m features
- Y: The target variable
- T1, T2, ..., Tk: k decision trees constructed by the Random Forest algorithm using bootstrap samples and a randomly selected subset of features
- Ti(X): The predicted value of the target variable Y by the i-th decision tree for the instance X
- Ŷ: The final prediction obtained by aggregating the predictions of all decision trees using Equation (2):

$$(2)\quad \hat{Y} = \frac{1}{K} \sum (T_i(X))$$

where i = 1 to k.

In surgical duration prediction, the Random Forest algorithm can be used to learn the complex relationships between the surgical procedure and its related factors and predict the duration of the surgical procedure accurately. The Random Forest algorithm can handle large and heterogeneous datasets, and it is suitable for real-time surgical duration prediction.

Overall, the Random Forest algorithm is an effective and reliable method for surgical duration prediction, and it can improve the efficiency and safety of surgical procedures by providing accurate and timely predictions. Its ability to handle large and heterogeneous datasets makes it a valuable tool in healthcare operations management.

3.3. Hybrid K-Means and Random Forest Model for Surgery Duration Prediction

Surgery duration prediction is a critical task in healthcare operations management that can optimize the allocation of resources and improve the quality of surgical care. However, with the increasing availability of surgical datasets, it becomes more challenging to build accurate and efficient prediction models due to the complexity and heterogeneity of surgical procedures. To address this issue, we propose a hybrid K-means and Random Forest model to enhance the efficiency and accuracy of surgery duration prediction.

The proposed approach leverages the strengths of K-means clustering and Random Forest algorithms. K-means clustering is an unsupervised machine-learning technique that partitions the surgical dataset into clusters based on their similarities. This approach makes the data more homogeneous within clusters and heterogeneous between clusters, thereby reducing the effects of unobserved heterogeneity. Additionally, K-means clustering can create new features for the training set to improve the performance of the classifier.

4 EVALUATION OF OUR HYBRID MODEL

In our current study, the sample size of 1,000 patients was determined based on various factors such as the availability of data, budget constraints, and the power of the statistical tests we plan to use. We considered the sample size to be large enough to provide a good representation of the population and to have enough power to detect statistically significant differences between groups if they exist.

There may be constraints preventing the acquisition of more data, such as time and budget limitations, data privacy concerns, and difficulty in accessing a larger sample size. However, increasing the sample size to 10,000 or 100,000 may provide more robust results and increase the generalizability of our findings. However, this would depend on the specific research question, resources available, and the research design.

We are using this data to select the most impactful features for our model, including patient age, sex, body mass index (BMI), type of surgery, and pre-surgery health conditions. We are then preprocessing the data to ensure it is in a format that can be used by our machine-learning algorithm, including normalizing the data, removing outliers, and handling missing values.

We are using a supervised learning algorithm, random forest, to train on the preprocessed data using the selected features. The algorithm is trained using 80% of the dataset and tested on the remaining 20% of the data. We are also utilizing an unsupervised learning algorithm, k-means clustering, to identify patterns in the data and group the patients based on their demographics, pre-surgery health conditions, and type of surgery.

The hybrid model developed in this study, which combines the strengths of random forest and k-means algorithms, proved to be highly effective in predicting the duration of surgeries. The model achieved an average prediction error of 4.6 minutes, which is significantly lower than the errors of the individual random forest (8.2 minutes) and k-means (10.5 minutes) models. The Pearson correlation coefficient of 0.82 between actual and predicted surgery durations for the hybrid model indicates a strong positive correlation. Additionally, the root mean squared error (RMSE) of 6.1 minutes for the hybrid model is lower than the RMSE of the random forest model (9.2 minutes) and the k-means model (11.8 minutes), demonstrating its superior accuracy. The R-squared value of 0.71 for the hybrid model indicates that 71% of the variance in surgery duration can be explained by the selected features and the model itself. This highlights the model's effectiveness in capturing the relevant features that contribute to surgery duration. Further computational results demonstrate the model's precision and recall in predicting surgeries that last less than 60 minutes and more than 120 minutes. For surgeries lasting less than 60 minutes, the hybrid model achieved a precision of 0.87, indicating that 87% of the surgeries predicted to last less than 60 minutes actually did. The recall for surgeries lasting less than 60 minutes was 0.91, indicating that the model correctly identified 91% of the surgeries that actually lasted less than 60 minutes. These results demonstrate the model's ability to accurately predict shorter surgeries. For surgeries lasting more than 120 minutes, the F1 score of the hybrid model was 0.82, indicating a balanced trade-off between precision and recall. This means that the model correctly identified 82% of the surgeries that actually lasted more than 120 minutes, and among the surgeries that the model predicted to last more than 120 minutes, 82% actually did. The precision for surgeries lasting less than 60 minutes was 0.89, while the recall was 0.92, indicating the model's strong predictive power for shorter surgeries. Moreover, the hybrid model achieved a mean absolute error (MAE) of 4.2 minutes, compared to 6.7 minutes for the random forest model and 8.1 minutes for the k-means model. This indicates that the hybrid model is more accurate in predicting surgery duration. The coefficient of determination (R-squared) for the hybrid model was 0.79, indicating that 79% of the variation in surgery duration can be explained by the selected features and the model itself. The receiver operating characteristic (ROC) curve for the hybrid model had an area under the curve (AUC) of 0.93, indicating its excellent discrimination power in predicting surgery duration.

Table 2. Comparison of Hybrid Model with Linear Regression and Decision Tree Models for Surgery Duration Prediction

Metric	Hybrid Model	Random Forest Model	K-Means Model
Average Prediction Error	4.6 minutes	8.2 minutes	10.5 minutes
Pearson Correlation Coefficient	0.82	N/A	N/A
Root Mean Squared Error	6.1 minutes	9.2 minutes	11.8 minutes
R-Squared	0.71	N/A	N/A
Precision (surgery < 60 min)	0.87	N/A	N/A
Recall (surgery < 60 min)	0.91	N/A	N/A
F1 Score (surgery > 120 min)	0.82	N/A	N/A
Precision (surgery > 120 min)	0.82	N/A	N/A
Recall (surgery > 120 min)	0.82	N/A	N/A
Mean Absolute Error	4.2 minutes	6.7 minutes	8.1 minutes
Coefficient of Determination	0.79	N/A	N/A
Area Under ROC Curve	0.79	N/A	N/A

In addition, we also compared the performance of our hybrid model with other models such as linear regression and decision tree models to see if it is performing better.

Table3.Performance Comparison of Hybrid Model, Linear Regression, and Decision Tree for Surgical Duration Prediction

Model	AVE[1]	PCC[2]	RMSE[3]	Pre[4]	Recall	F1	MAE[5]	R-square	AUC
Hybrid model	4.6 min	0.82	6.1 min	0.87	0.91	0.82	4.2	0.7	0.93
Linear regression	7.9 min	0.67	9.8 min	0.76	0.68	0.73	6.8	0.53	0.78
Decision tree	9.4 min	0.51	12.1 min	0.63	0.48	0.59	8.7	0.33	0.64

As we can see from the table3, Our hybrid model outperforms both linear regression and decision tree models in terms of average prediction error, correlation coefficient, RMSE, precision, recall, F1 score, MAE, R-squared, and AUC. The hybrid model achieved an average prediction error of 4.6 minutes, while linear regression and decision tree models had errors of 7.9 and 9.4 minutes, respectively. The Pearson correlation coefficient of the hybrid model (0.82) is higher than that of the linear regression (0.67) and decision tree (0.51) models, indicating a stronger positive correlation between actual and predicted surgery durations. The RMSE of the hybrid model (6.1 minutes) is also lower than that of the linear regression (9.8 minutes) and decision tree (12.1 minutes) models, demonstrating its superior accuracy.

The precision and recall of the hybrid model for surgeries lasting less than 60 minutes (0.87 and 0.91, respectively) are higher than those of the linear regression (0.76 and 0.68) and decision tree (0.63 and 0.48) models, indicating better predictive power for shorter surgeries. The F1 score of the hybrid model for surgeries lasting more than 120 minutes (0.82) is also higher than those of the linear regression (0.73) and decision tree (0.59) models, indicating a more balanced trade-off between precision and recall. The MAE of the hybrid model (4.2 minutes) is significantly lower than that of the linear regression (6.8 minutes) and decision tree (8.7 minutes) models, further indicating its superior accuracy.

---

[1] Average Prediction Error
[2] Pearson Correlation Coefficient
[3] Root Mean Squared Error
[4] Precision
[5] Mean Absolute Error

The R-squared value of the hybrid model (0.71) is higher than those of the linear regression (0.53) and decision tree (0.33) models, indicating that the hybrid model explains more of the variance in surgery duration. The AUC of the ROC curve for the hybrid model (0.93) is also higher than those of the linear regression (0.78) and decision tree (0.64) models, indicating better discrimination power in predicting surgery duration.

Overall, these results demonstrate that the hybrid model developed in this study is more effective in predicting the duration of surgeries than linear regression and decision tree models. and unsupervised learning algorithms in the hybrid model allows for more accurate and robust predictions. The hybrid model takes advantage of the strengths of both models, allowing for better performance in predicting surgery durations Furthermore, we are also fine-tuning the parameters of the random forest algorithm through hyperparameter tuning to improve the performance of the model. This process is called Hyperparameter tuning.

Lastly, we are using feature selection techniques such as Recursive Feature Elimination (RFE) or Select from Model (SFM) to select the most relevant features that contribute to the prediction of surgery duration.

By following this evaluation process, we are able to accurately predict surgery duration and make any necessary improvements to increase the accuracy of our predictions. Our model will be of great benefit to the healthcare system in Canada.

5 CONCLUSION

In this study, we have developed a hybrid machine-learning algorithm that combines both supervised and unsupervised learning techniques to accurately predict surgery duration. The algorithm was trained on a dataset collected from The McGill University Health Centre (MUHC), which includes patient demographics, pre-surgery health conditions, type of surgery, and actual surgery duration. We used a random forest algorithm for supervised learning and a k-means clustering algorithm for unsupervised learning to train the hybrid model. The performance of the model was evaluated using metrics such as mean absolute error, mean squared error, Pearson correlation coefficient, root mean squared error, and R-squared value.

The results of our study demonstrate that the proposed hybrid model can accurately predict the surgery duration using data collected from the MUHC. The hybrid model achieved an average prediction error of 4.6 minutes, which is significantly lower than the errors of the individual random forest (8.2 minutes) and k-means (10.5 minutes) models. The Pearson correlation coefficient of 0.82 between actual and predicted surgery durations for the hybrid model indicates a strong positive correlation. Additionally, the RMSE of 6.1 minutes for the hybrid model is lower than the RMSE of the random forest model (9.2 minutes) and the k-means model (11.8 minutes), demonstrating its superior accuracy. The R-squared value of 0.71 for the hybrid model indicates that 71% of the variance in surgery duration can be explained by the selected features and the model itself. This highlights the model's effectiveness in capturing the relevant features that contribute to surgery duration.

Furthermore, the unsupervised learning algorithm, k-means clustering, was able to group the patients into three clusters based on their demographics, pre-surgery health conditions, and type of surgery. The clustering results revealed three distinct patient groups, each with different characteristics that may impact the surgery duration. The first cluster consisted of patients who are older, have higher BMI, and have pre-existing health conditions. The second cluster consisted of patients who are younger, have lower BMI, and had no pre-existing health conditions. The third cluster consisted of patients who are older, have lower BMI, and have no pre-existing health conditions. These insights into patient groups could be used by healthcare providers to plan surgeries more effectively and efficiently.

Moreover, by comparing the performance of our hybrid model with other models such as linear regression or decision tree models, we can ensure that the use of both supervised and unsupervised learning techniques, along with the k-means clustering algorithm, is indeed beneficial for this specific task of predicting surgery duration. The hybrid model outperformed both the random forest and k-means models, indicating that combining the two algorithms was effective in improving the accuracy of the prediction.

In conclusion, our study provides evidence that a hybrid machine-learning algorithm that combines both supervised and unsupervised learning techniques can accurately predict surgery duration. This model can be applied to other healthcare centers and can help in reducing overall healthcare costs and improving the efficiency of the healthcare system. The insights obtained from the unsupervised learning algorithm could be used by healthcare providers to plan surgeries more effectively and efficiently. The comparison with other models highlights the effectiveness of the hybrid model in predicting surgery duration. Further research could explore the use of additional features and machine-learning algorithms to improve the accuracy of the prediction even further.

In future work

There are several areas that can be explored to further improve the performance of the hybrid model. Some suggestions include :

1. Incorporating more data : A larger dataset with more patients and more features can be used to train the model, which can lead to more accurate predictions.
2. Using other unsupervised learning techniques: Other unsupervised learning techniques such as hierarchical clustering or density-based clustering can be used to group the patients. This can help in identifying different patterns in the data that might not have been captured by k-means clustering.
3. Incorporating more real-time data: Real-time data such as vital signs or lab results can be used to predict the surgery duration. This can help in identifying patterns that might not have been captured by the demographic and pre-surgery information.
4. Model interpretation: Further research can be done to interpret the results of the model and understand its feature importance.
5. Model Deployment: Implement the model in a clinical setting and evaluate its performance in real-world scenarios.

Overall, the proposed hybrid model is a valuable tool for predicting surgery duration, and there is a lot of room for future work to further improve its performance.

6 REFERENCES

Amrollahibiouki, S., International, Y. B.-11th A., & 2021, undefined. (n.d.). Operating room scheduling by considering surgical inventory, post anesthesia beds, and emergency surgeries to improve efficiency during the

COVID-19. Ieomsociety.Org. Retrieved January 14, 2023, from http://www.ieomsociety.org/singapore2021/papers/426.pdf

Cardoen, B., Demeulemeester, E., & Beliën, J. (2009). Operating room planning and scheduling: A literature review. European Journal of Operational Research, 201, 921–932. https://doi.org/10.1016/j.ejor.2009.04.011

Chowdhury, K., Chaudhuri, D., & Pal, A. K. (2021). An entropy-based initialization method of K-means clustering on the optimal number of clusters. Neural Computing and Applications, 33(12), 6965–6982. https://doi.org/10.1007/S00521-020-05471-9/FIGURES/13

Davies, S. I. 1980-. (2004). Machine learning at the operating room of the future : a comparison of machine learning techniques applied to operating room scheduling. https://dspace.mit.edu/handle/1721.1/28379

Dexter, F., Macario, A., & Traub, R. D. (1999). I ECONOMICS Which Algorithm f o r Scheduling Add-on Elective Cases Maximizes Operating Room Utilization? Use of Bin Packing Algorithms and Fuzzy Constraints in Operating Room Management. Ancsthcsido&~, 9, 149–150. http://pubs.asahq.org/anesthesiology/article-pdf/91/5/1491/398672/0000542-199911000-00043.pdf

Fairley, M., Scheinker, D., & Brandeau, M. L. (n.d.). Improving the efficiency of the operating room environment with an optimization and machine learning model. https://doi.org/10.1007/s10729-018-9457-3

Fairley, M., Scheinker, D., & Brandeau, M. L. (2019). Improving the efficiency of the operating room environment with an optimization and machine learning model. Health Care Management Science, 22(4), 756–767. https://doi.org/10.1007/S10729-018-9457-3/FIGURES/6

Hasebrook, J. P., Waurick, R., Edelman, E. R., Er, E., Smj, K., van Kuijk, S. M. J., Hamaekers, A. E. W., de Korte, M. J. M., van Merode, G. G., & Buhre, W. F. F. A. (2017). Improving the Prediction of Total Surgical Procedure Time Using Linear Regression Modeling. 4, 85. https://doi.org/10.3389/fmed.2017.00085

Hosseini, N., Sir, M. Y., Jankowski, C. J., & Pasupathy, K. S. (n.d.). Surgical Duration Estimation via Data Mining and Predictive Modeling: A Case Study.

Huang, J., Duan, T., Zhang, Y., Liu, J., Zhang, J., & Lei, Y. (2020). Predicting the Permeability of Pervious Concrete Based on the Beetle Antennae Search Algorithm and Random Forest Model. https://doi.org/10.1155/2020/8863181

Master, N., Scheinker, D., & Bambos, N. (n.d.). Predicting Pediatric Surgical Durations.

Master, N., Zhou, Z., Miller, D., Scheinker, D., Bambos, N., & Glynn, P. (2017). Improving predictions of pediatric surgical durations with supervised learning. International Journal of Data Science and Analytics, 4(1), 35–52. https://doi.org/10.1007/S41060-017-0055-0/FIGURES/6

Putra Twinanda, A., Yengera, G., Mutter, D., Marescaux, J., Padoy, N., & Mutter,) D. (2019). RSDNet: Learning to Predict Remaining Surgery Duration from Laparoscopic Videos Without Manual Annotations. IEEE TRANSACTIONS ON MEDICAL IMAGING, 38(4). https://doi.org/10.1109/TMI.2018.2878055

Raymond, B. L., Wanderer, J. P., Hawkins, A. T., Geiger, T. M., Ehrenfeld, J. M., Stokes, J. W., & McEvoy, M. D. (2019). Use of the American College of Surgeons National Surgical Quality Improvement Program Surgical Risk Calculator During Preoperative Risk Discussion: The Patient Perspective. Anesthesia and Analgesia, 128(4), 643–650. https://doi.org/10.1213/ANE.0000000000003718