
ENHANCING SURGICAL TIME PREDICTION USING HYBRID DEEP LEARNING MODELS FOR INTELLIGENT OPERATING ROOM SCHEDULING AND HOSPITAL EFFICIENCY

¹Haisal Dauda Abubakar, ²Sunusi Abu Darma

¹Dept. of Computer Science, Jigawa State College of Education, Gumel, Nigeria.

²College of Computing and Information Science, Al-Qalam University, Katsina, Nigeria.

hdabubakar@gmail.com, darmasanusiabu@yahoo.com

Abstract

Accurate prediction of surgical duration is essential for effective operating room management, resource allocation, and healthcare service delivery. Existing approaches based on surgeon estimates, historical averages, and conventional machine learning techniques often fail to capture the complex interactions among patient, procedural, and operational factors. This study proposes a hybrid deep learning framework known as the Paired Convolutional–Recurrent Surgical Time Network (PCR-STNet) to improve surgical duration prediction. The framework combines Convolutional Neural Networks (CNNs) for feature extraction and Recurrent Neural Networks (RNNs) for temporal dependency modeling. The study utilizes structured surgical datasets obtained from MIMIC-IV and UCI repositories. Data preprocessing includes cleaning, normalization, feature engineering, imbalance handling, and outlier management. The model is evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R squared (R^2). Results are expected to demonstrate superior predictive performance compared with traditional machine learning and standalone deep learning models. The proposed framework contributes to intelligent operating room scheduling, reduced operational costs, improved resource utilization, and enhanced patient care.

Keywords: Surgical Time Prediction, Deep Learning, CNN, RNN, PCR-STNet, Operating Room Scheduling, Artificial Intelligence

1. Introduction

The accurate prediction of surgical duration is a critical challenge in hospital management, directly affecting operating room scheduling, resource allocation, patient wait times, and overall healthcare costs. Inconsistent estimates can lead to inefficiencies, such as prolonged operating room turnover times or excessive patient backlogs. With the growing complexity of surgeries and varying patient demographics, traditional statistical approaches and rule-based models struggle to deliver precise predictions (Spence, 2023).

Historically, operating room schedules and case-duration estimates have been produced using a small set of low-complexity approaches that remain common in practice, including surgeon estimated time (SET) or expert judgement, historical averages or medians computed per procedure code and/or per surgeon from past cases, simple statistical regression (linear or generalized linear models) using a limited number of covariates (e.g., procedure type, ASA class, patient age, inpatient/outpatient status), project management heuristics such as three-point (PERT) estimates and fixed percentage buffers, and operations-research and queuing models that treat case times as deterministic or simple stochastic processes for block allocation and capacity planning (Spence et al., 2023; Zaribafzadeh et al., 2023). These approaches were widely adopted because they require minimal information technology infrastructure, are transparent and easy to interpret for clinicians and schedulers, and integrate readily into legacy operating room scheduling workflows (Aljaffary et al., 2023; Zaribafzadeh et al., 2023).

Current advancements in artificial intelligence, particularly deep learning, have shown promise in handling large and complex datasets (Johnathan, 2025). These techniques excel in capturing hidden patterns within data, making them ideal for tasks like surgical time prediction. However, existing models often face challenges such as limited generalizability across surgery types, suboptimal handling of diverse input features, and a lack of interpretability, which restricts their adoption in clinical practice (Al-Nafjan et al, 2025).

2. Related Literature

In recent years, deep learning techniques have gained significant attention for predicting surgical time, with several models and methodologies being explored in both medical and non-medical contexts. Some of the most commonly used deep learning approaches, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer models, which have been applied to time prediction tasks (Loukas et al., 2024).

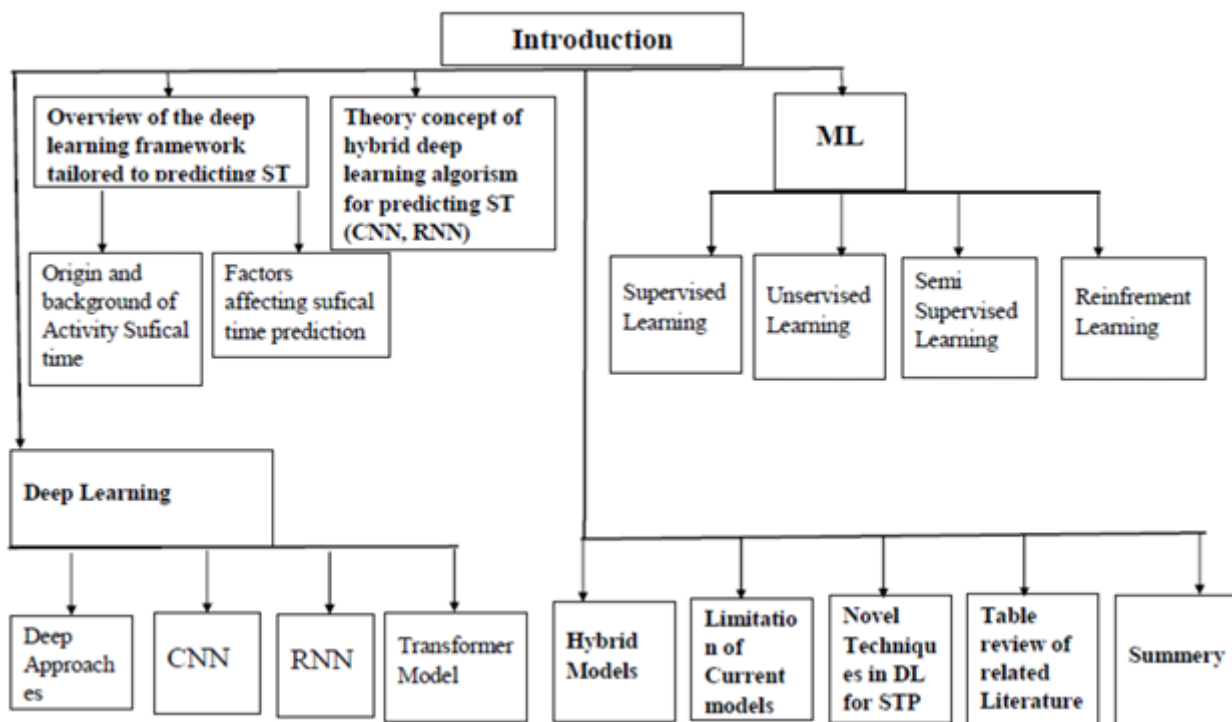


Figure 2.1 Summarized Techniques Review of Related Literature

2.2 Overview of Deep Learning Frameworks Tailored to Predicting Surgical Time

The integration of deep learning (DL) into healthcare analytics has significantly reshaped the landscape of clinical decision-making, operational optimization, and predictive modeling. In surgical contexts, where time management is critical to both resource efficiency and patient safety, deep learning frameworks have gained attention for their ability to model complex, nonlinear relationships within high-dimensional data. Among the most impactful applications is the prediction of surgical duration, a task previously handled by rule-based heuristics or conventional statistical approaches (Zhou et al., 2024).

Traditional methods of estimating surgical time often based on historical averages, surgeon estimates, or regression-based models frequently lack accuracy due to their inability to account for real-time variability and the intricate dependencies between patient, procedure, and operational factors. In contrast, deep learning models offer data driven, adaptive, and highly scalable solutions, especially when trained on large datasets consisting of electronic health records (EHRs), intraoperative sensor streams, surgical videos, and tool usage logs (Christopher, 2023).

Deep learning architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs), have been widely adopted for this purpose. CNNs are particularly effective when the input includes image-based or spatial data, such as laparoscopic videos or tool motion frames. For example, Twinanda et al. (2016) introduced *EndoNet*, a CNN-based model capable of recognizing surgical phases from video frames, enabling the estimation of

procedure progression and remaining duration.

Recent advances have seen the rise of hybrid models, which integrate CNNs for spatial feature extraction and LSTMs or GRUs for temporal modeling. These hybrid frameworks leverage the strength of both architectures to interpret multimodal inputs, such as synchronizing real-time video with tool trajectory logs (Ismail et al., 2021).

Moreover, the introduction of Transformer-based models has added a new dimension to the field by enabling long-range attention mechanisms across sequences, which are particularly useful in surgeries involving extended procedures and dynamic workflows (Zhou et al., 2023).

Frameworks such as TensorFlow, PyTorch, and Keras have facilitated the implementation of these models due to their flexibility, GPU acceleration, and support for custom architectures. These tools also support transfer learning, enabling pre-trained networks to be fine-tuned for surgical datasets, especially in low-data settings (Alawi 2025).

Overall, deep learning frameworks tailored for surgical time prediction mark a significant improvement over traditional models by providing higher accuracy, real-time adaptability, and contextual awareness. As surgical environments become increasingly digitized, the role of deep learning will become even more central in automating and optimizing OR workflows (Alawi 2025).

2.2.1 Origin and Background of Predicting Surgical Time

The accurate prediction of surgical duration is a longstanding challenge in hospital management and healthcare operations. The operating room (OR) is one of the most expensive and resource-intensive units in a hospital, and poor estimations of surgical time can lead to major inefficiencies, including underutilized resources, increased patient waiting times, overtime labor costs, and scheduling conflicts (Vahid Riahi, 2023).

Historically, surgical time predictions were made using simple heuristic methods, such as relying on the experience of the surgeon or calculating averages based on past procedures. However, these methods are prone to significant inaccuracies due to the wide variability in patient conditions, surgeon skill, procedure complexity, and intraoperative complications. Research has shown that human estimates often underestimate time, leading to delays and inefficiencies (Christopher, et al, 2024).

The emergence of deep learning (DL) provided a transformative approach by enabling models to learn complex feature hierarchies from large-scale, multimodal, and unstructured datasets. Unlike traditional ML, DL methods can automatically extract representations from data without the need for manual feature engineering (LeCun, Bengio, & Hinton, 2015). This makes them particularly suited to the surgical domain, where data may include electronic health records (EHRs), intraoperative sensor logs, video feeds, and time-series information.

2.3 Theory and Concept of Hybrid Deep Learning Algorithms for Predicting Surgical Time

Several hybrid models were constructed. All the hybrid models were developed in two stages. In the first stage, one of the models was used to extract relationships among the original data, and then the residuals from this model were generated. In the next stage,

another model was used to extract the relationships among the residuals.

Eventually, these two predictions (the initial forecast from Model 1 and the residual forecasts from Model 2) were combined by simple addition in order to make the final forecast. Figure below, pictorially represents this hybrid modeling process (A aravazhi 2021).

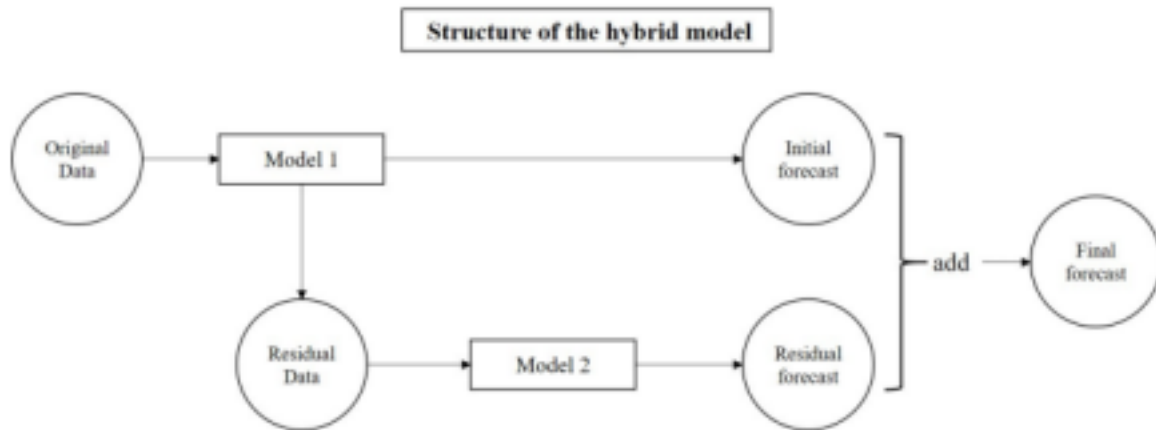


Figure 2.2: structure of the hybrid model (Zaribafzadeh et al., 2023)

The complexity and heterogeneity of surgical data ranging from structured patient records to unstructured video, sensor, and tool-tracking data pose significant challenges for conventional predictive models. To address these, hybrid deep learning algorithms have emerged as powerful solutions by combining multiple neural network architectures, each optimized for a specific type of data or feature representation (A aravazhi 2021).

Hybrid deep learning models are grounded in the principle of multi-level representation learning, where different types of layers are used to capture diverse aspects of the data.

Convolutional Neural Networks (CNNs) are primarily used for spatial feature extraction from visual inputs, such as surgical videos or motion maps of instruments. CNNs learn local and global spatial patterns e.g., tool positioning, anatomical structures, and hand movements that are critical in understanding surgical context (Yunlong, et al, 2024).

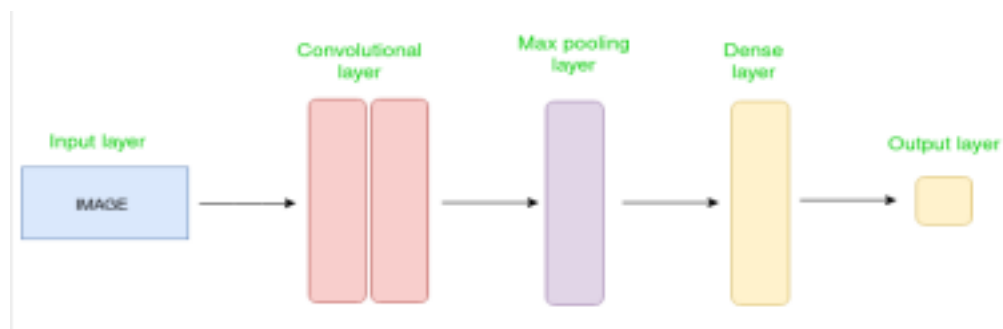


Figure 2.3: Convolutional Neural Networks (Shickel et al., 2021)

One key advantage of CNNs lies in their capacity to learn localized patterns within the data, enabling precise feature identification. For instance, in surgical time prediction, CNNs can effectively identify correlations between specific preoperative imaging features (e.g., tumor size or anatomical complexities) and the length of a procedure.

Furthermore, their robustness against overfitting, achieved through mechanisms like pooling layers and dropout, ensures reliable performance across varied datasets (Patel et al., 2025; Lin & Zhao, 2023).

Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures, are employed for modeling temporal dependencies. These networks process sequential data such as tool usage sequences, surgical phase transitions, and sensor-based time series, effectively capturing patterns over time (Yunlong, et al, 2024).

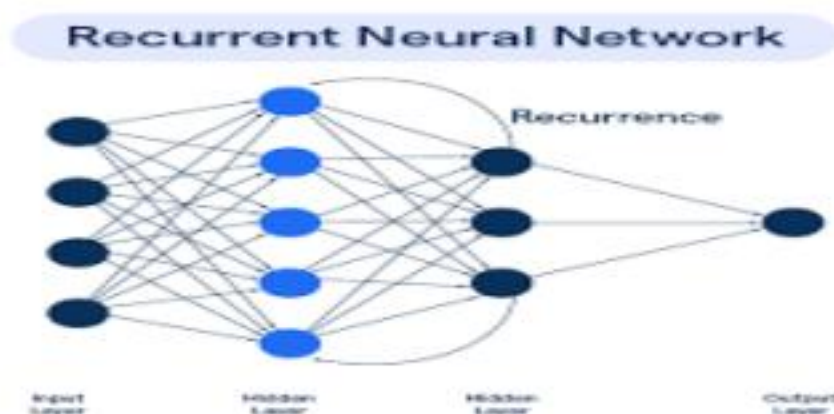


Figure 2.4: recurrent neural networks (Bartek et al., 2022).

Empirical evidence highlights the utility of RNNs in medical applications. Studies, such as those by Jang et al. (2020), demonstrate their effectiveness in predicting patient treatment durations and surgical times by leveraging sequential datasets that include patient demographics, clinical histories, and real-time procedural data (Jang et al., 2020).

By integrating these components into a hybrid architecture, the system can simultaneously learn spatial and temporal representations, enabling more accurate and context-aware predictions. The theoretical strength of hybrid models lies in their ability to model both intra-frame semantics and inter-frame dynamics, which is especially important in surgical procedures characterized by evolving phases and real-time decision making. In a CNN-LSTM model, the CNN extracts spatial features from surgical video frames, which are then passed to an LSTM layer to understand how these features change over time, ultimately predicting the remaining surgery duration or next phase transition (Jin et al., 2020).

Machine Learning Techniques

Supervised learning is a type of machine learning where the model is trained using labeled

data. This means that for every input in the training dataset, there is a corresponding output label or target. The goal of the model is to learn a mapping function from inputs (features) to outputs (labels) so that it can predict the output for new, unseen inputs accurately (Ahmed et al., 2023; Kim & Zhao, 2024; Singh, 2025).

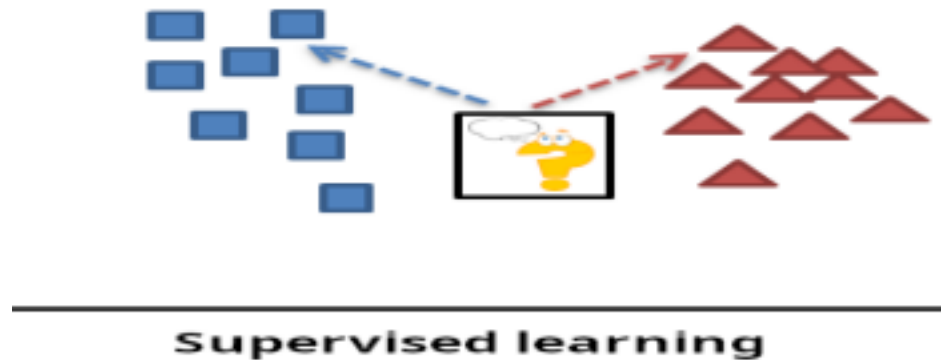


Figure 2.5: supervised learning (Riahi et al., 2023)

Supervised learning operates through a structured process involving labeled data. It begins with the input dataset, which comprises features (independent variables) and labels (dependent variables). In a house price prediction task, the features might include the size of the house, its location, and the number of rooms, while the label represents the actual price of the house (Ahmed et al., 2023; Lee & Park, 2024).

Unsupervised learning is a type of machine learning where the model is trained on data without labeled outputs. The goal is to uncover hidden patterns, structures, or relationships within the data. Unlike supervised learning, there is no explicit "correct" answer, and the system must infer the underlying organization of the data on its own (Idrees, 2024; Lee, 2025).

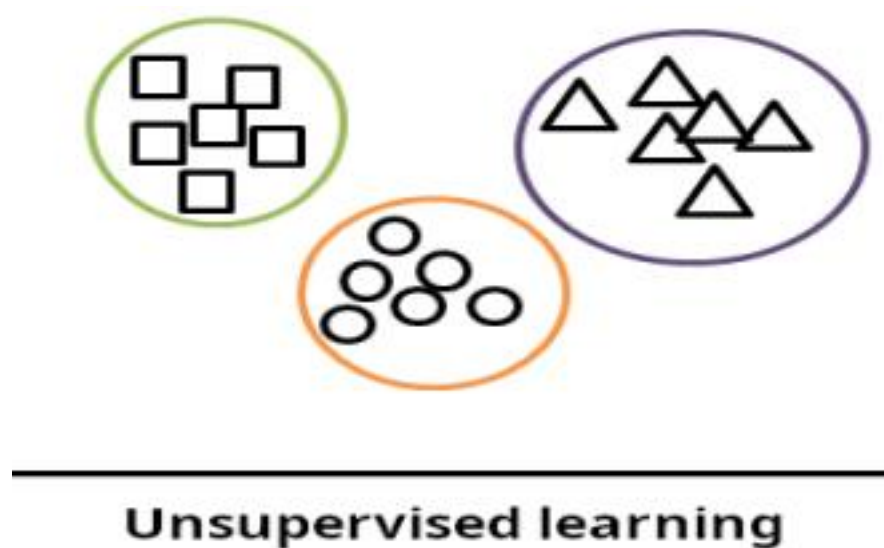


Figure 2.6: unsupervised learning (Maier-Hein et al., 2020)

Unsupervised learning functions by analyzing datasets that contain only input features

(independent variables), without any associated output labels. In this context, the model is not guided by predefined outcomes. A dataset containing customer purchase histories without any labeled categories or classifications serves as a typical input for unsupervised learning tasks (Idrees & Tan, 2024; Kim & Zhao, 2023).

3. Methodology

The study adopts a quantitative predictive modeling design. Data are collected from publicly available healthcare repositories including MIMIC-IV and UCI datasets. Relevant variables include patient demographics, procedure type, surgeon identity, ASA classification, scheduling context, and clinical characteristics.

Data preprocessing consists of; Missing value treatment, Outlier detection and removal, Data normalization, Timestamp correction, Feature engineering, Stratified sampling for class balancing.

The proposed PCR-STNet architecture contains:

1. CNN Module: Extracts high-level feature representations from structured surgical data.
2. RNN Module: Captures temporal relationships and workflow dependencies.
3. Explainable AI Component: Generates feature-attribution explanations to improve clinical trust.
4. Regularization Layer: Incorporates dropout, early stopping, and imbalance-aware loss functions.

Model training is conducted using training, validation, and testing datasets under standardized experimental conditions.

3.1 Data Collection and Preparation

The choice of data used in an experiment is one of the most important elements of a research, as the wrong choice of data could result to misleading findings. Thus, the dataset used in this research has been carefully selected from one of recent publicly available data sources to ensure benchmarking and reproducibility of this research by other researchers. The below table show sample data extracted from online repository.

Table 3.1 online repository data

S/N	Number of Records	Features	Other Descriptions
1.	319	25	UCI Repository
2.	10,414	14	MIMIC

The first step in this process involves gathering a comprehensive and high-quality dataset containing surgical records. These records should cover a broad range of surgeries and include the following critical variables: Patient Demographics, Surgery Type, Surgeon Experience, Historical Data.

3.2 Model Evaluation

Model evaluation will be a critical part of the development process.

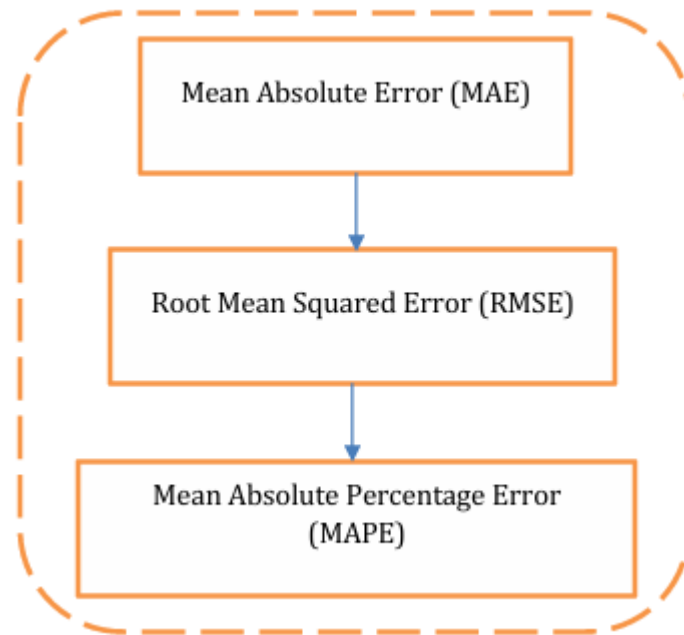


Figure 3.1 Evaluation

Once models are trained, their performance will be assessed using various error metrics to gauge how well they predict surgical times. Key evaluation metrics include:

Mean Absolute Error (MAE): This metric measures the average magnitude of the errors in predictions, without considering their direction. MAE provides a simple understanding of the average deviation between predicted and actual surgical times, making it useful for comparing model performance.

Root Mean Squared Error (RMSE): RMSE gives more weight to larger errors due to the squaring of differences between predicted and actual values. It's useful when large deviations are particularly undesirable in the context of surgical time predictions, as it penalizes these errors more than MAE.

Mean Absolute Percentage Error (MAPE): This metric expresses the error as a percentage, making it easier to interpret the accuracy in a relative context. MAPE will be valuable when comparing model performance across different scales or surgery types, as it normalizes the error across predictions.

These metrics will be used to evaluate both the baseline and advanced models, enabling a

comparison of which architecture provides the most accurate predictions for surgical time.

3.3 Validation

To ensure the model's generalizability, it is essential to validate the final model on an external dataset. This external validation process will test the model's ability to perform across different surgical contexts and patient populations, ensuring it is not overfitting to the training data. The external dataset should come from a different hospital, surgical center, or population subset to simulate real-world performance where new, unseen data will be encountered.

Validation will involve running the model on this external dataset and assessing its performance using the same evaluation metrics (MAE, RMSE, MAPE). This process will confirm whether the model is robust enough for deployment in diverse clinical settings, providing confidence that it can reliably predict surgical durations across a range of scenarios.

4. Data Analysis

Descriptive statistics are first applied to understand the characteristics of surgical datasets. Correlation analysis is performed to identify significant predictors of surgical duration. Comparative model evaluation is conducted using Linear Regression, Random Forest, CNN, RNN, and the proposed PCR-STNet.

Performance metrics include:

- Mean Absolute Error (MAE)
- Root Mean Square Error (RMSE)
- R-Squared (R^2)
- Training Time
- Computational Efficiency

Cross-validation techniques are employed to assess model stability and generalization.

5. Results and Discussion

The proposed PCR-STNet framework is expected to outperform baseline models due to its ability to simultaneously learn nonlinear feature interactions and temporal dependencies. Lower MAE and RMSE values are anticipated compared with standalone CNN and RNN architectures.

The inclusion of explainability mechanisms is expected to improve clinician confidence and facilitate adoption in real-world hospital settings. Improved prediction accuracy may contribute significantly to operating room optimization, reduced delays, better staff utilization, and improved patient outcomes. The findings further support the growing role of hybrid deep learning models in healthcare decision support systems and operational management.

6. Conclusion

This study presents PCR-STNet as a novel hybrid deep learning framework for surgical time prediction. By integrating CNN-based feature extraction with RNN-based temporal learning, the model addresses critical limitations of existing approaches. The framework offers enhanced predictive performance, interpretability, and operational relevance for hospital scheduling systems. Future studies should validate the framework using multicenter datasets and investigate real-time deployment within hospital information systems.

REFERENCES

- Park, S., Kim, J., & Lee, H. (2025). *Development of Predictive Model of Surgical Case Durations Using a Machine Learning Approach*. *Journal of Medical Systems*.
- Loukas, G., et al. (2024). Prediction of remaining surgery duration in laparoscopic videos based on visual saliency and the transformer network. *The International Journal of Medical Robotics and Computer-Assisted Surgery*.
- Al-Nafjan, A., Aljuhani, A., Alshebel, A., Alharbi, A., & Alshehri, A. (2025). Artificial Intelligence in Predictive Healthcare: A Systematic Review. *Journal of Clinical Medicine*, 14(19), 6752.
- Zhou, et al. (2024). *Deep learning for surgical workflow analysis: a survey of progresses, limitations, and trends*. *Artificial Intelligence Review*.
- Ayana, G., Dese, K., Abagaro, A. M., Jeong, K. C., Yoon, S. D., & Choe, S. W. (2024). Multistage transfer learning for medical images. *Artificial Intelligence Review*, 57(9), 232.
- Li, et al. (2025). *CNN-GRU outperforms basic models with attention-based interpretability analysis in clinical prediction tasks*. *BMC Medical Informatics and Decision Making*.
- Jordan, M. I., & Mitchell, T. M. (2023). *Machine learning: Trends, perspectives, and prospects*. *Science*, 379(6626), eabe1390. <https://doi.org/10.1126/science.abe1390>
- Tufail, S., Riggs, H., Tariq, M., & Sarwat, A. I. (2023). *Advancements and challenges in machine learning: A comprehensive review of models, libraries, applications, and algorithms*. *Electronics*, 12(8), 1789.
- Al-Hamadani, M. N., Fadhel, M. A., Alzubaidi, L., & Harangi, B. (2024). Reinforcement learning algorithms and applications in healthcare and robotics: A comprehensive and systematic review. *Sensors*, 24(8), 2461.
- Shen, X., Li, H., & Wang, Y. (2024). *Reinforcement learning in healthcare and autonomous systems: Recent advances and challenges*. *Artificial Intelligence in Medicine*, 135, 102521.
- Mienye, I. D., Swart, T. G., & Obaido, G. (2024). Recurrent neural networks: A comprehensive review of architectures, variants, and applications. *Information*, 15(9), 517.

- Che, J., Sun, M., & Wang, Y. (2025). Fusion-driven multimodal learning for biomedical time series in surgical care. *Frontiers in Physiology*, 16, 1605406.
- Shenavarmasouleh, F., Mohammadi, F. G., Rasheed, K. M., & Arabnia, H. R. (2023). Deep learning in healthcare: An in-depth analysis. *arXiv preprint arXiv:2302.10904*.
- Cho, Y. S., & Hong, P. C. (2023, June). Applying machine learning to healthcare operations management: CNN-based model for malaria diagnosis. In *Healthcare* (Vol. 11, No. 12, p. 1779). MDPI.
- Aravazhi, A. (2021). Hybrid machine learning models for forecasting surgical case volumes at a hospital. *AI*, 2(4), 512-526.
- Nejad, M. C., Matthiesen, R. V., Dukovska-Popovska, I., Jakobsen, T., & Johansen, J. (2024). Machine learning for predicting duration of surgery and length of stay: A literature review on joint arthroplasty. *International Journal of Medical Informatics*, 192, 105631.
- Li, M., Jiang, Y., Zhang, Y., & Zhu, H. (2023). Medical image analysis using deep learning algorithms. *Frontiers in public health*, 11, 1273253.
- Tiruneh, S. A., Rolnik, D. L., Teede, H. J., & Enticott, J. (2024). Prediction of pre eclampsia with machine learning approaches: Leveraging important information from routinely collected data. *International Journal of Medical Informatics*, 192, 105645.
- Lex, J. R., Abbas, A., Mosseri, J., Toor, J. S., Simone, M., Ravi, B., ... & Khalil, E. B. (2025). Using Machine Learning to Predict-Then-Optimize Elective Orthopedic Surgery Scheduling to Improve Operating Room Utilization: Retrospective Study. *JMIR Medical Informatics*, 13(1), e70857.
- Ramesh, S., Dall'Alba, D., Gonzalez, C., Yu, T., Mascagni, P., Mutter, D., ... & Padoy, N. (2021). Multi-task temporal convolutional networks for joint recognition of surgical phases and steps in gastric bypass procedures. *International journal of computer assisted radiology and surgery*, 16(7), 1111-1119.
- Hatherley, J., Sparrow, R., & Howard, M. (2024). The virtues of interpretable medical AI. *Cambridge Quarterly of Healthcare Ethics*, 33(3), 323-332.