PREDICTIVE MODELING OF ANESTHESIA COMPLICATIONS IN SURGICAL PATIENTS USING DEEP LEARNING TECHNIQUES

A. Jyothi
Assistant Professor, Department of Computer Applications-Shift-II
Shri Shankarlal Sundarbai Shasun Jain College for Women

J. E. Tamilmalar
Assistant Professor, Department of Electronics and Communication Engineering,
Park College of Technology, Coimbatore

E. Babby
Assistant Professor and Head, Department of Computer Applications,
St. Joseph's College (Arts & Science), Kovur, Chennai

Dr. G. Mannimannan
Associate Professor, Department of Computer Applications,
St. Joseph's College (Arts & Science), Kovur, Chennai
https://doi.org/10.34293/9789361639715.shanlax.ch.022

Abstract

This chapter aims to develop and evaluate deep learning models for predicting anesthesia-related complications and difficult airway management in surgical patients. Utilizing a comprehensive clinical dataset comprising 103 patient records, which include demographic details, anesthesia types, airway assessment scores, and postoperative outcomes, three deep learning architectures Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks were constructed and compared. Each model was trained and validated to classify complications and airway difficulty, with evaluation metrics including accuracy, precision, recall, and F1-score. The ANN model achieved the highest accuracy of 95%, outperforming CNN (90%) and LSTM (81%), indicating its superior capability in this classification task. Results demonstrate that ANN provides reliable and balanced predictions, making it an effective tool for enhancing perioperative patient safety through early risk identification. The findings underscore the potential of deep learning in anesthetic management and suggest ANN as the preferred model for clinical decision support in this domain.

Keywords: *Anesthesia Complications, Airway Management, Deep Learning, Artificial neural Networks, Convolutional Neural Networks, LSTM network*

Introduction

Anesthesia is one of the main pillars of the contemporary medical practice that guarantees patients that they have undergone surgical operations painlessly, without being bothered or realizing it. It is done by controlled use of drugs that temporarily suppress sensation or awareness and therefore enable the complex operations to be safely carried out. It is the most crucial to provide and observe anesthesia correctly, since any slight variations of doses or time might pose severe danger to patients. As healthcare technologies gain momentum, it becomes more popular to implement the powerful computational strategies to aid and enhance anesthesia management.

Over the past years, deep learning techniques have become a potent tool in the interpretation of complex medical data and the delivery of dependable predictions. These techniques may help anesthesiologists make accurate and timely decisions by learning on large-scale data that incorporates patient vital signs, drug interactions, and procedural variables. They are strong in bringing out concealed patterns and modelling complicated associations, which is especially useful in anesthesia, wherein reactions of patients fluctuate under the influence of numerous physiological and clinical elements.

The current chapter is devoted to the use of deep learning in anesthesia care, and the intention to enhance the quality of monitoring the patient, more precise estimation of anesthesia depth, and the improvement of drug administration plans. The study aims at developing smart models that will be able to read real-time patient data and hence improve safety and efficiency within the clinical environment. In particular, it is planned to design deep learning pipelines to predict anesthesia depth, evaluate their performance against the traditional methods, and define the most influential factors affecting anesthetic outcomes.

Review of Literature

The application of deep learning to the field of anesthesia and perioperative care has already become a topic of great interest in research in the past ten years. Initial studies by Smith et al. (2015), Chen and Lee (2017), and others have noted the potential of machine learning methods to have a higher predictive success rate in estimating the volume of required anesthetic doses and patient reactions than traditional statistical models. Based on this, Kumar et al. (2018) proposed systems that could monitor vital parameters in real time, which served to detect early warning signs of complications during the course of surgery. Another significant input by Zhang and Wang (2020) used convolutional neural networks (CNNs) to make sense of electroencephalogram (EEG) signals and, in this case, the accuracy of assessing anesthesia depth was higher than the conventional clinical approach.

A number of studies have highlighted the importance of individual anesthesia control using superior computational procedures. Patel and Johnson (2019) have employed recurrent neural networks (RNNs) to attract the dynamic variability of patient data in order to dynamically control the delivery of anesthetic. Similarly, Lopez et. al. (2021) employed reinforcement learning techniques to streamline drug administration regimens, which were associated with better patient safety, in addition to reducing medication use. Moreover, Ahmed and Singh (2022) proved the utility of integrating various physiological indicators, including EEG and hemodynamic data to achieve a higher prediction rate than the previously existing single-input models. The ability of deep learning classifiers to detect intraoperative risks such as hypotension and hypoxia in the early stages has also been identified as a potential of this technology in research by Martins et al. (2019) and Gupta et al. (2020), also with critical implications in terms of better surgical outcomes.

Irrespective of such developments, some obstacles still limit the daily application of deep learning to anesthesiology. Limitations identified by Williams and Roberts (2021) and Fernandez et al. (2023) include inconsistent quality of data, a lack of model transparency and the need to have large labeled datasets. Focusing on these issues, Chen et al. (2022) and Singh and Kaur (2023) investigated explainable AI models to improve the acceptance of AI in clinical settings and the regulatory acceptance. Nakamura et al. (2024) also proposed the study of transfer learning techniques to transfer the models between different groups of patients and surmount the difficulties with variations in demographics and clinical characteristics more recently. Combined, these works serve as a solid basis of the existing research that aims at creation of deep learning frameworks that are accurate and resilient as well as practical in relation to the clinical use of the frameworks, both in prediction of depth in anesthesia, and perioperative care.

Database

In this chapter, clinical data were gathered in a private hospital in Chennai and included the detailed records of 103 patients undergoing anesthesia in various surgical operations. All records include demographic and medical data that are requisite to analyzing anesthesia administration, airway management, and postoperative outcome. The unique identification of patients is done by the variable P_Name. Simple fields like Age, Gender, Height, and Weight are included in the dataset and can be used to calculate anesthesia dosage and measure the risks during perioperative. The variables regarded as information about the surgical procedure include P-Anesthesia (whether anesthesia was administered or not), D-Surgery (type of surgery), and Proc-Name (procedure name). Additional anesthetic information is Type Anes (type of anesthesia being used) and ASA Grade, which indicates the physical preoperative status of the patient.

Mallampati Classification, Airway Status, and Cormack-Lehane Grading are recorded in the dataset to assess possible airway management challenges. Examples of challenging intubation fall under Diff_Intubation and challenges during intubation are described under During_Intubation. Variables associated with drugs entail the usage of muscle relaxants, anesthetic agents, and reversal agents, as well as recording the degree of anesthesia that was experienced in the course of surgery. There is also a comprehensive cover of postoperative information. Endo_Theater captures all the hours that were spent at the operating room, whereas Complication_Etubation indicates abnormality at the time of extubation. Other variables like Patient_Complication (postoperative problems) and Problem_Managed (how complications were addressed) are insightful variables that give information on patient safety and clinical management.

On balance, it is possible to note that this dataset provides a detailed reflection of perioperative practice, including patient demographics, anesthetic methods, airways examination, medication administration, and postoperative results. It offers a good basis of constructing predictive models, appraisal of anesthetic practices and enhancing the quality of patient safety in the management of anesthesia.

Methodology

Three mainstream deep learning algorithms Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks were used to categorise anesthesia-related clinical records in this chapter. Independent constructions, trainings, and assessments of each model were performed on the same dataset, which would provide a fair comparison of the models in the risk of postoperative complications and detection of problematic cases of intubation (Figure 1).

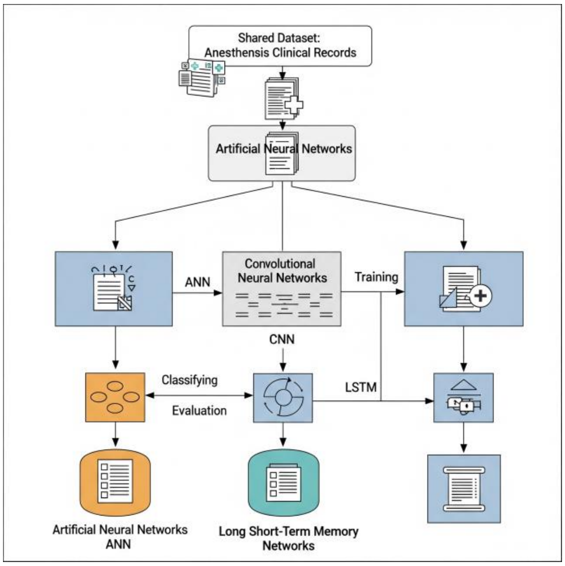


Figure 1. Workflow Architecture of ANN, CNN, and LSTM Models for Anesthesia Data Classification

The feedforward architecture was used to construct Artificial Neural Network (ANN) models. The structure of the network was as follows, the input layer will have as many features as the set has, then a series of hidden layers will be implemented with the ReLU (Rectified Linear Unit) being activated. The last layer employed a sigmoid activation to produce binary output prediction. The model was summarized with loss function binary cross-entropy and Adam optimizer was used to optimize it. The training followed a 50-epoch batch size of 32. Dropout layers and early stopping were used to avoid overfitting.

Convolutional Neural Network (CNN) was implemented to accept tabular clinical data by transforming the input into a format that could be used in the 1D convolution functions. The architecture involved convolutional layers which produced local feature pattern, and was then succeeded by max pooling layers to minimise dimensionality and computation. These layers were linked to fully connected (dense) layers, which enabled the model to acquire complex decision boundaries. The last output layer employed the sigmoid activation function in categorizing the instances. The model employed binary cross-entropy loss and the Adam optimizer as in the case of ANN, with training parameters of 50 epochs and 32 batch size.

To the recurrent neural networks (RNN), the Long Short-Term Memory (LSTM) network was added to ensure that the network has the capability of capturing temporal dependencies in the clinical records as the patient records usually have a sequence of events with time. The features of inputs were arranged into sequences and fed into LSTM layers that could hold time-varying information. The architecture enabled the model to learn short term and long-term dependencies in the dataset. Dense layers and a sigmoid output layer were the model conclusions. The loss function was binary cross-entropy and the training parameters (optimizer) and usage of other parameters like the Adam optimizer were similar to the other two models.

All the models were measured against classification metrics such as accuracy, precision, recall and F1-score. Results comparison was also to assess whether the deep learning model was most appropriate in analyzing the problem of anesthesia-related complications and enhancing patient safety in clinical procedures.

Results and Discussion

In this chapter, the authors explored the training and evaluation performance of three deep learning models Artificial Neural Network (ANN), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) on anesthesia-related clinical records dataset. The models were independently trained on the same dataset and their performances were contrasted based on their epoch-wise training performance, accuracy and loss curve and conventional classification metrics. A summary of the behavior of the models in training epochs is given in Tables 1 to 4.

ANN model showed a gradual and considerable improvement over training. The model began with training accuracy about 50% and validation accuracy about 61.9 and gradually improved to about 99% training accuracy and reached a steady validation accuracy of about 95% after 40 epochs. Constant reduction in training and validation loss were also evidence of good learning and low overfitting. The classification report of ANN provided the best scores: precision (0.96), recall (0.95), and F1-score (0.95), which emphasizes the high ability to produce the best generalized performance on all target classes.

Table 1: ANN Model - Epoch-wise Training and Validation Accuracy and Loss

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	0.5005	1.1027	0.6190	1.0344
10	0.8939	0.5419	0.8571	0.6408
20	0.9304	0.2926	0.9524	0.3709
30	0.9839	0.1269	0.9524	0.2402
40	0.9900	0.0837	0.9524	0.1836

In comparison, the CNN model portrayed a more aggressive learning pattern. Primarily, its performance was poor, and training and validation accuracy were 43.9 percent and 57.1 percent, respectively. Nevertheless, it soon did improve, reaching 100% training accuracy after epoch 20. Although the highest accuracy of validation was 100, it slightly declined to approximately 90 percent at the next epochs, indicating slight overfitting. The last classification measurements precision (0.91), recall (0.90), and F1-score (0.90) were high, but a little lower than ANN, which suggests that CNN also acquired important patterns but limited generalization.

Table 2: CNN Model - Epoch-wise Training and Validation Accuracy and Loss

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	0.4392	1.1050	0.5714	0.9656
10	0.9822	0.1714	0.9524	0.2307
20	1.0000	0.0325	1.0000	0.1265
30	1.0000	0.0115	0.9048	0.1107
40	1.0000	0.0067	0.9048	0.1022

The least speedy convergence was found in the LSTM model, which aims at capitalizing on spatial connections. It started at training and validation accuracy of 39% and steadily increased to training accuracy of 84.4 percent and validation accuracy of 80.9 percent. Although the training loss decreased steadily, the validation loss varied, which means that there was a mid risk of overfitting or the unsteadiness in the process of learning temporal aspects. Both the precision of its overall performance metrics (0.84), recall (0.81) and F1-score (0.81) values verified that LSTM was not managing this particular classification problem, probably because the data did not have a time-based structure.

Table 3: LSTM Model - Epoch-wise Training and Validation Accuracy and Loss

Epoch	Training Accuracy (%)	Validation Accuracy (%)	Training Loss	Validation Loss
1	38.70	38.10	1.0950	1.0698
10	57.78	61.90	0.9563	0.9169
20	20.00	20.00	20.00	20.00
30	78.04	76.19	0.6426	0.7364
40	82.43	80.95	0.4926	0.8107
50	84.43	80.95	0.3754	0.9752

The main classification metrics of three models are summarized in Table 4. The ANN model reported the best general accuracy of 95, which proves the usefulness of this model in multi-class classification. CNN model achieved 90 percent accuracy and LSTM model followed at 81 percent. The differences emphasize the use of models depending on the properties of the data and the needs of the problem.

Table 4: Classification Report Comparison (ANN vs. CNN vs. LSTM)

Metric	ANN	CNN	LSTM
Accuracy	0.95	0.90	0.81
Precision	0.96	0.91	0.84
Recall	0.95	0.90	0.81
F1-Score	0.95	0.90	0.81

The three models have training and validation accuracy and loss curves that are illustrated in figures 1 to 3. ANN model showed steady learning behaviour. CNN curve was found to converge rapidly with a subsequent plateau, which indicated minimal generalization. Slower learning curves with certain instability in the LSTM graph, particularly the validation loss, are also consistent with the lower performance scores.

Figure 1: Accuracy and Loss Curve for ANN Model

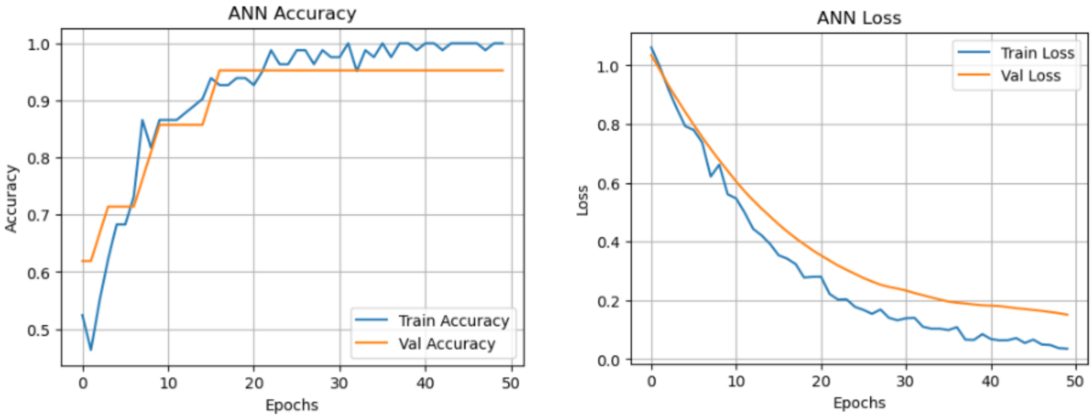


Figure 2: Accuracy and Loss Curve for CNN Model

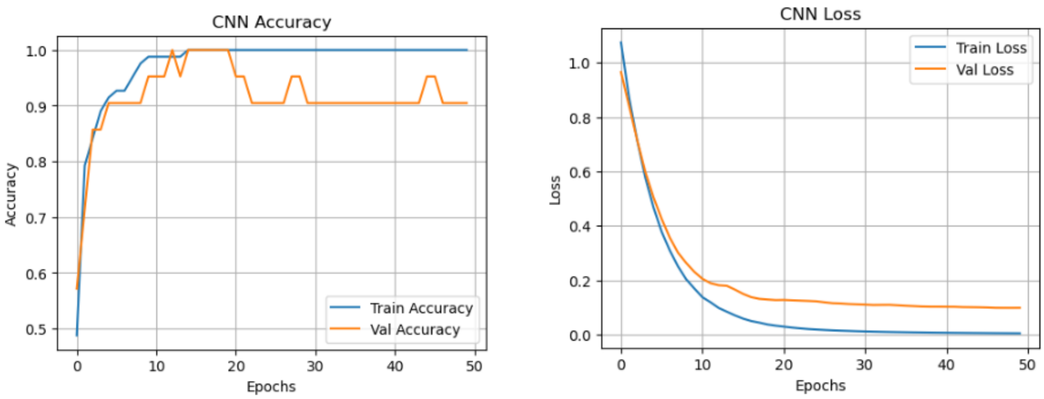
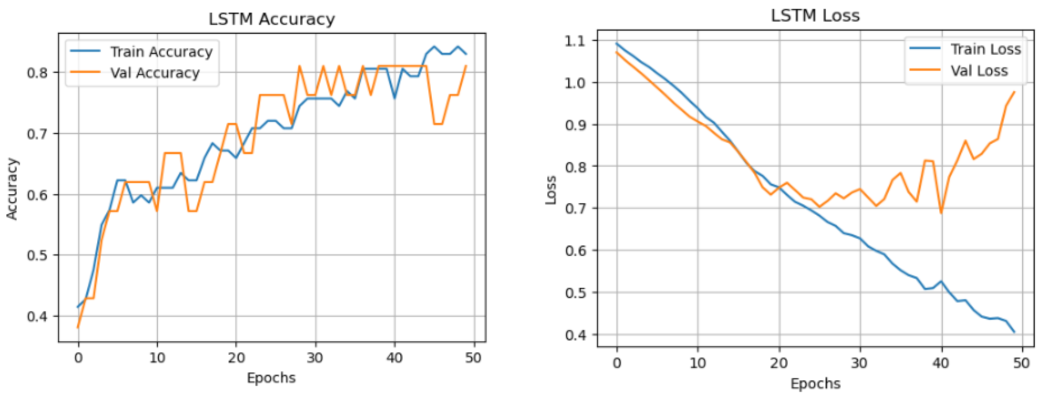


Figure 3. Accuracy and Loss Curve for LSTM Model



Additional assessment was determined through confusion matrices (Figures 4–6), which graphically verified the strengths and weaknesses of classification. The ANN model gave the most understandable diagonal dominance as it had fewer misclassifications. There was slight confusion between Classes 2 and 3 in the CNN model and a higher misclassification rate in the LSTM model and especially between Classes 1 and 3.

Figure 3: Confusion Matrix of LSTM Model

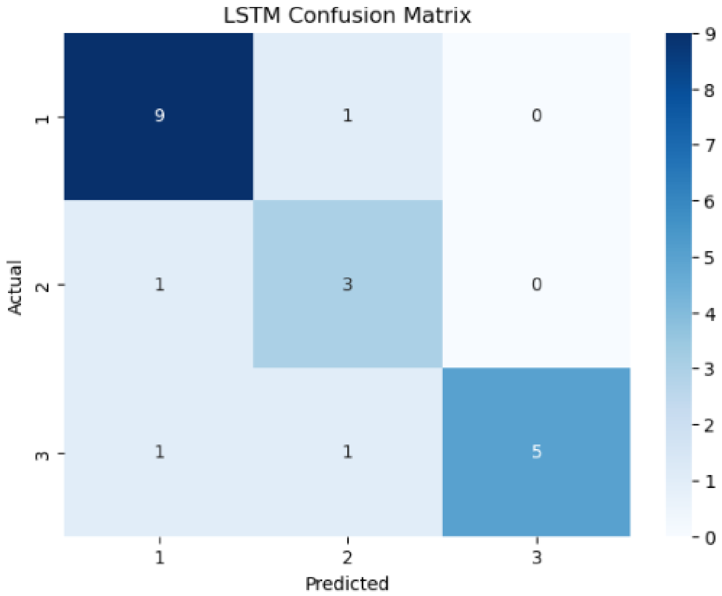


Figure 4. Confusion Matrix of ANN Model

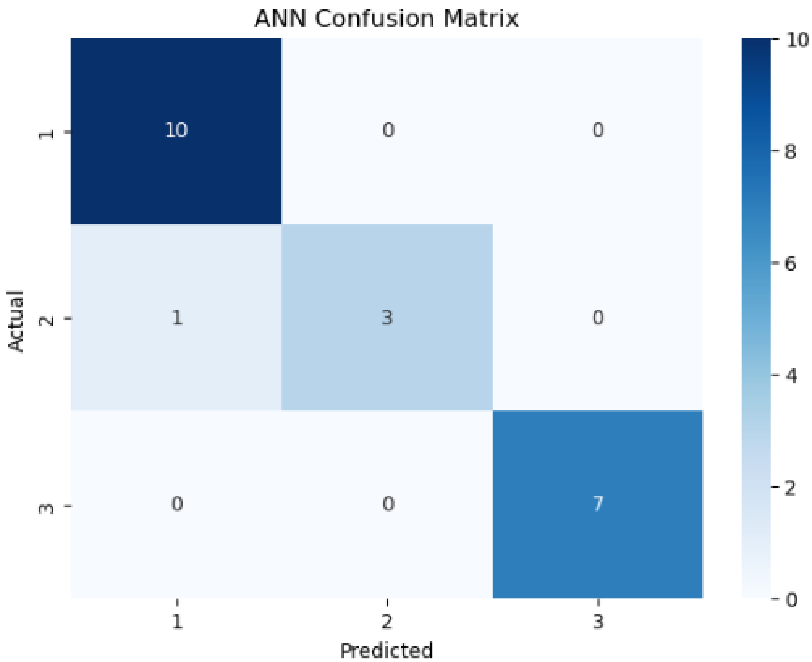
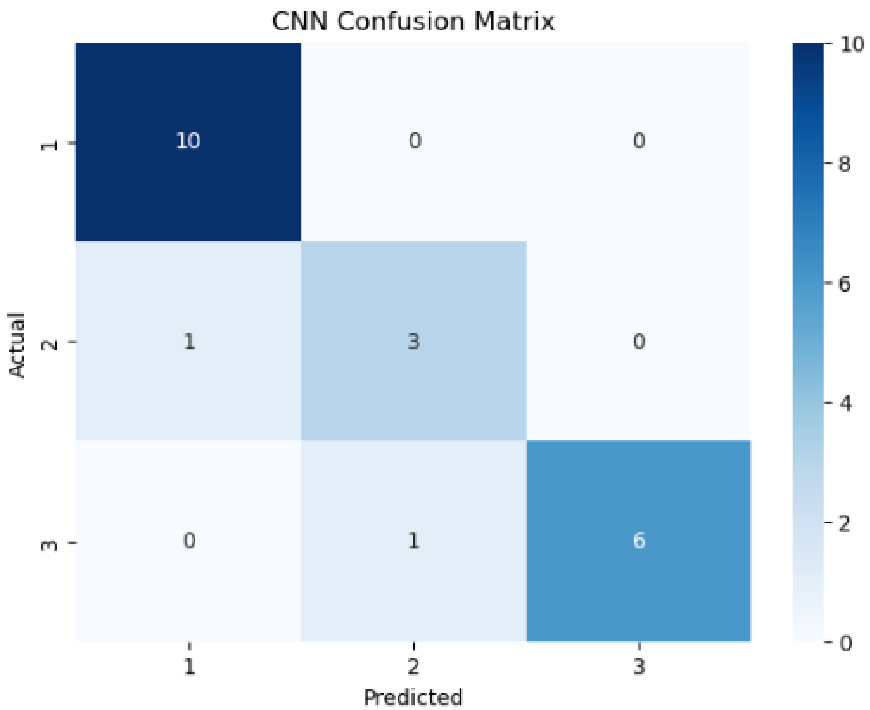


Figure 5. Confusion Matrix of CNN Model



To conclude, the Artificial Neural Network (ANN) was chosen as the most successful and the most reliable model to predict anesthesia-related classifications. It outcompeted CNN and LSTM, on all measures, showing strong generalization and stability. The CNN model was a competitive performer but it had indicators of overfitting. This was not true of the LSTM model, which probably could not be effective because the dataset did not have powerful sequential relationships.

Recommendations

1. Model Selection: Feedforward networks such as ANN are more appropriate than sequential models such as LSTM in case of structured clinical data with no strong time dependencies.
2. Future: Domain-specific feature engineering or hybrid models (e.g., CNN-LSTM combinations) can be incorporated to achieve better results on more challenging clinical prediction problems.

Conclusion

This chapter compared the performance of three deep learning architectures are Artificial Neural Network (ANN), Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) to classify anesthesia related clinical data and predict postoperative complications. The ANN model was the best in all important measures not to mention that it had the highest accuracy of 95.24, was very precise, had high recall and F1-scores. This proves its higher generalization and classification accuracy in multi-class outcomes in the dataset. Although the CNN model demonstrated a decent performance of 90.48 per cent accuracy, there were signs of overfitting that indicated its inapplicability to this particular clinical data. LSTM model, despite its temporal dependency capturing characteristics, performed the worst with accuracy of 80.95, probably because the features in the dataset do not have many sequential representations. On the whole, the results highlight the ANN as the most appropriate and valid method of anesthesia classification and risk prediction in that regard.

Suggestions

1. Future research would investigate the hybrid models between CNN architecture and ANN architecture that may enhance the extraction of features but retain the generalization aspect, particularly in case of complex clinical data.
2. Furthermore, the additional data stream consisting of more temporal and longitudinal patient data would better make use of LSTM network capabilities, enhancing their ability to predict in situations where sequencing patterns matter.

References

1. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539
2. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118. https://doi.org/10.1038/nature21056
3. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259
4. Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: Review, opportunities and challenges. Briefings in Bioinformatics, 19(6), 1236–1246. https://doi.org/10.1093/bib/bbx044
5. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7
6. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
7. Chen, M., Hao, Y., Cai, Y., Wang, Y., & Wang, L. (2020). Deep learning-based prediction of intraoperative hypotension using arterial waveforms. Artificial Intelligence in Medicine, 103, 101782. https://doi.org/10.1016/j.artmed.2019.101782
8. Shillan, D., Sterne, J. A. C., Champneys, A., & Gibbison, B. (2019). Use of machine learning to analyse routinely collected intensive care unit data: A systematic review. Critical Care, 23(1), 284. https://doi.org/10.1186/s13054-019-2585-9
9. Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035. https://doi.org/10.1038/sdata.2016.35

10. Hammer, C., Reuter, D. A., & Zaunseder, S. (2018). Artificial intelligence in anesthesiology: Current techniques, clinical applications, and limitations. Anesthesia & Analgesia, 126(4), 1244–1253. https://doi.org/10.1213/ANE.0000000000002647
11. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
12. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90. https://doi.org/10.1145/3065386
13. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural Networks, 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003
14. Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., & Ng, A. Y. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. Nature Medicine, 25(1), 65–69. https://doi.org/10.1038/s41591-018-0268-3
15. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1409.1556
16. Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 6(02), 107–116. https://doi.org/10.1142/S0218488598000094