International Journal of Medical and All Body Health Research

Multimodal Deep Learning for Intraoperative Hemodynamic Instability Prediction

Dr. Vivek Vaibhav [1*], Dr Niyati Sinha [2], Dr Nikhil vaid [3], Dr. Megha Maheshwari [4], Dr. Manish madhrey [5], Dr. Shail Ashokbhai Patel [6]

[1] Head, Department of Anesthesia and Critical Care, India

[2-6] Department of Anesthesia, Saraswathi Institute of Medical Science, Hapur, India

* Corresponding Author: **Dr. Vivek Vaibhav**

Article Info

E-ISSN: 2582-8940
ISSN (online): 2582-8940
Volume: 07
Issue: 02
Received: 20-02-2026
Accepted: 18-03-2026
Published: 10-04-2026
Page No: 45-58

Abstract

One significant cause of perioperative morbidity and mortality is intraoperative hemodynamic instability (IHI), which includes episodes of hypotension, hypertension, and arrhythmias. Current monitoring systems mainly alert clinicians when thresholds are violated, relying heavily on clinician alertness; therefore, they are reactive in nature and do not have sufficient sensitivity to detect the range of complex, multidimensional physiological trajectories that occur before IHI. This article will present a systematic review of the emerging multimodal deep learning (MDL) approaches to predicting IHI intraoperatively by using different sources of data to evaluate how heterogeneous continuous physiological waveforms (e.g., arterial blood pressure, ECG, SpO2, EEG), electronic health records (EHRs), anesthesia information management systems (AIMS), and imaging modalities can be fused and analyzed through early, late, or hybrid fusion architectures. We will examine neural networks, including convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and transformer-based models, to determine their effectiveness for capturing both spatial and temporal signal characteristics relevant to the prediction of hemodynamics. Studies of apparent groundbreaking importance such as Hypotension Prediction Index (HPI) studies and VitalDB-based studies have shown that MDL models yield area under the receiver operating characteristic curve (AUROC) values between 0.91 and 0.95, along with clinically relevant lead times (5-15 minutes) before hemodynamic deterioration; resulting in significant improvements over standard statistical and univariate machine-learning baselines. Clinical studies to evaluate the efficacy of the MDL approach for clinician decision support have shown that the use of MDL-based decision support will reduce time-weighted average intraoperative hypotension by up to 50% relative to control groups in randomized controlled trials. Additionally, this review provides an overview of significant barriers (i.e., data heterogeneity, model generalizability, model bias, regulatory compliance, and ethical use) that must be overcome for successful implementation of an AI system in high-stakes operating rooms. Future directions for MDL systems include implementation of federated learning frameworks, the development of personalized adaptive predictive systems, and the utilization of explainable AI (XAI) to promote clinician confidence and enhance the equitable and safe use of MDL systems in diverse perioperative care environments.

DOI: https://doi.org/10.54660/IJMBHR.2026.7.2.45-58

Keywords: Multimodal Deep Learning, Intraoperative Hemodynamic Instability, Hypotension Prediction, Physiological Waveform Analysis, Anesthesia Decision Support, Multimodal Data Fusion, Explainable AI in Perioperative Care

Introduction

1. Background and Clinical Significance of Intraoperative Hemodynamic Instability

1.1. Definition and Classification of Intraoperative Hemodynamic Instability

Hemodynamic instability in patients during surgery is a general term that can mean any problem with heart function that occurs during the time a patient is undergoing surgery (the perioperative period) where there are no problems with heart function from a normal healthy state (homeostasis) [1, 2].

The main presentation of this instability is intraoperative hypotension, which is usually defined as a mean arterial pressure (MAP) below 65 mmHg that occurs for at least one to five minutes; however, this definition may differ among studies [3]. In addition to intraoperative hypotension, intraoperative hypertension (MAP >105 mmHg or blood pressure >160 mmHg) tachycardia, bradycardia, and complex arrhythmias such as atrial fibrillation can also occur de novo or can be an exacerbation of an existing arrhythmia due to surgery [4].

The hemodynamic status of a patient undergoing surgery under general or regional anesthesia is influenced by an ever-changing array of factors (anesthetic agents: propofol, volatile agents, opioids; mechanical ventilation; fluid status; surgical stimulation; autonomic tone) that work in a temporally rich and non-linear fashion to create a physiological signal space that cannot be characterized accurately through point observations or threshold monitoring. Therefore, hemodynamic deterioration events often go unrecognized by conventional alarm systems, resulting in delays in recognition of these events and providing suboptimal treatment [6].

1.2. Clinical Risks and Outcomes Associated with Hemodynamic Instability

IHI leads to many important clinical problems. Most of the research has focused on sustained intraoperative hypotension, which has been shown to be causally associated with MINS, AKI, POCD, stroke, and increased perioperative mortality [7, 8]. A landmark study including more than 16,000 patients undergoing certain types of non-cardiac surgeries found that there was a dose-dependent relationship between time spent with MAP less than 65 mmHg and 30-day mortality; every minute spent below this threshold indicates an increase in the likelihood of having an organ injury [9]. Cardiac surgery patients experience even greater consequences from their time spent with inadequate blood pressure control due to a compromised cardiopulmonary reserve and the effects of hemodynamic instability associated with cardiopulmonary bypass [10]. The potential dangers associated with intraoperative hypertension, although less well researched, are similar, and an increase in the likelihood of having significant cerebrovascular accidents, aortic dissections, and myocardial ischemia occurs, especially for patients with pre-existing hypertensive heart disease [11]. Additionally, arrhythmias that may occur during the operative procedure, such as ventricular tachycardia, complete heart block, or severe bradycardia, are likely to lead to hemodynamic collapse resulting in an emergency resuscitation. The total impact that all of these events have on negative patient outcomes, the consumption of available healthcare resources, and liability will only increase the need to implement more predictive than reactive management of hemodynamic status in the perioperative period.

1.3. Limitations of Traditional Monitoring and Prediction Methods

The ASA Basic Monitoring Standards mandate that standard intraoperative monitoring includes ECG (electrocardiogram), pulse oximetry, non-invasive blood pressure, capnography (i.e., end-tidal) and temperature monitoring all continuously during the procedure [12]. When all these monitoring modalities are used together to help assess the patient's cardiovascular and respiratory stability, they can provide a very comprehensive picture of the physiological status of the patient. However, each of these monitoring modalities is reactive, which means that they will alert the anesthesia provider once there has been some deterioration in the patient's cardiovascular or respiratory status. With the time between the actual physiological deteriorating event and the alarm activation, and high false-positive rates for traditional alarms using a threshold-based system are two major contributing factors for alarm fatigue in providers [13].

Risk stratification algorithms prior to surgery, such as the ASA-PS physical status classification, the revised cardiac risk index, and the Lee index provide static estimates of the perioperative risk of a patient prior to surgery. However, these algorithms do not provide a sufficient level of temporal resolution or integrative power to be able to assist a provider in predicting intraoperative hemodynamic events with clinical utility [14]. Machine learning algorithms using structured preoperative data alone yield only marginally improved risk stratification compared to traditional risk algorithms. These machine learning algorithms still do not provide the level of clinical utility due to their inability to incorporate high-frequency, dynamic physiologic data found only in the intraoperative period [15].

1.4. The Emerging Need for Predictive Analytics in Perioperative Care

A major shift has occurred in the use of predictive analytics for perioperative care due to a combination of factors that include an explosion in computing power, widespread use of high-res bedside monitors, and the development of deep learning techniques that can analyze large quantities of varied data [16]. Predictive anesthesiology, a new way to integrate pre-surgical patient information, currently available real-time patient measurements, and drug information to provide prognostic assessments of whether a patient will experience hemodynamic instability, has gone from being a conceptual idea to becoming a prototype being used by anesthesiologists [17].

This change fits in with the growing movement toward personalized medicine, as there are likely more individual patient-specific developments than there are general trends in patient history. The anaesthetist now has the possibility of using a clinically validated decision support tool called MDL to predict whether hemodynamic instability will occur during surgery from 5 to 15 minutes into the future and therefore gives the anaesthetist an opportunity to adjust anaesthetic dosages, start using vasopressor medications, provide fluid replacement therapy, and communicate with the surgical team to turn a potential bad occurrence into a manageable process [18].

2. Fundamentals of Multimodal Deep Learning

2.1. Overview of Deep Learning Architectures

Deep learning is an advanced form of machine learning that uses multi-layered artificial neural networks to automatically learn progressively abstract and relevant features from raw or minimally pre-processed input data [19]. Unlike traditional machine learning processes which rely on manual feature engineering, deep neural networks develop optimal representations through an end-to-end gradient-based training process. This advantage has been particularly valuable in the clinical field because complex physiological signals often contain high dimensionality, non-linearity, and an incomplete understanding of relevant features [20].

Able to detect local patterns such as ECG waveform morphology and variations in arterial blood pressure (beat-to-beat) through the application of the learned filters through the spatial or temporal dimensions of the input signals (convolutional neural networks - CNNs). In contrast to CNNs, recurrent neural networks (RNNs) and their more stable alternative, long short-term memory networks (LSTMs) are specifically designed to model sequencing data by maintaining hidden states that propagate temporal context across time steps; thus making them very effective for modeling longitudinal physiological time series [21]. The transformer architecture [22] proposed by Vaswani *et al*. and then transforming natural language processing employs a multi-head self-attention mechanism that makes possible to model long-range dependencies across the entire sequence at any given time (i.e., without having the constraint of sequential processing where RNN's require) [22]. In fact, as it relates to the application of clinical time-series data, there is evidence of great promise as it has been adapted with positional encodings designed specifically for irregularly sampled clinical measurements.

2.2. The Concept and Rationale for Multimodal Learning

Multimodal Learning aims to develop a model that can handle various types of data (or data that will eventually come to it via multiple modes of delivery) with very different statistical properties and different sampling/rendering rates, temporal scales, and content values [23]. The potential clinical need for multimodal methods is directly related to the reality of biological complexity, where hemodynamic instability does not occur from the action of a single descriptor of a measurable quantity; instead, it arises from the simultaneous integration of all of the different factors affecting patient physiology, perioperative pharmacology, surgical context, and environment as represented across distinct streams of data [24].

To illustrate this, an example would be the case of an individual patient being predisposed to episodes of intraoperative hypotension that are influenced by a multitude of independent factors in real-time (e.g., arterial waveform characteristics). Still, they will also be affected by their baseline blood pressure measurements obtained before surgery, their medications, particularly diuretics/antihypertensives, fluid/blood products administered prior to surgery, surgical stage/positioning, anesthetic depth, and their autonomic reactivity—all of which are data elements appearing in four distinct data streams: electrocardiogram, arterial pressure waveform, electronic health record, and anesthesia administration log [25]. Any single data stream will impose an upper limit on model accuracy regardless of how sophisticated the model architecture may be. Consequently, the implementation of multimodal data integration not only represents a technical upgrade but also provides fundamental epistemological consistency with the multifactorial nature of physiologic perioperative hemodynamics.

2.3. Data Fusion Strategies: Early, Late, and Hybrid Fusion

Multimodal prediction tasks in clinical settings have been examined and developed using three main data fusion frameworks [26]. Early fusion occurs at the input level, where raw data from each of the multiple modalities is combined through concatenation or other methods before training the model. This allows us to learn representations that relate to the modal data at the time model training starts. This fusion type uses large amounts of computational resources and creates inter-modal correlations prior to feature training. Therefore, modality imbalance can create problems with high-frequency signal modalities taking over representation while lower dimensional EHR features are ignored or not properly represented [27].

Late fusion involves training independent unimodal models in parallel before combining their predictions, either through ensemble averaging, learned weights, or meta-classification. Late fusion maintains uniqueness for each modality but does not capture complementary interactions that occur only when the data from multiple modalities are processed simultaneously. Hybrid fusion, which is the most advanced fusion method, consists of modality-specific encoders with either cross-modal attention or gating mechanisms. These hybrid architectures utilize context-based relevance of each modality's features when implementing cross-modal integration, and implement multiple modalities into feature-fusing layers that can access the learning representations produced by the merged models [28]. Recent advancements in multimodal transformers with cross-attention modules and gated multimodal units (GMUs) provide improved performance over traditional late and hybrid fused models for complex prediction tasks, such as dynamic weighting of each modality's contribution based on fine-grained feature extraction techniques.

2.4. Advantages of Multimodal Approaches over Unimodal Models

The comparative advantages of multimodal deep learning versus unimodal models for perioperative prediction are well established in the literature [29]. Specifically, multimodal models are capable of producing higher AUROC values, improved calibration, and greater robustness to individual signal noise, missing data streams, and all sources of uncertainty through the use of redundant and complementary information across modalities. The incorporation of EHR-derived static patient features along with dynamic physiologic waveforms has been shown to be especially beneficial in improving early-stage prediction windows where subtle changes in the physiologic signal may be present but where the patient’s baseline risk is informative [30].

In addition, multimodal models are able to generalize better than single-modality models across all patient subpopulations, including the aged, those with chronic cardiovascular disease, and those patients undergoing high-acuity procedures, for whom single-modality models would typically demonstrate significant performance reductions. The capacity of multimodal models to allow the learner to selectively attend to the most diagnostic data stream at any given time enables the integration of all aspects of expert anesthesiologists' cognitive processes through the use of cross-modal attention mechanisms inherent in transformer-based multimodal architectures—this represents the mechanistic basis for the clinical utility of MDL systems.

3. Data Sources, Modalities, and Preprocessing

A comprehensive summary of multimodal data sources, their characteristics, and clinical relevance is provided in Table 1.

Table 1: Summary of multimodal data sources, associated monitoring systems, sampling characteristics, extracted features, and clinical relevance to intraoperative hemodynamic prediction.

Data Modality	Source System	Sampling Rate	Key Features	Clinical Relevance
Arterial Blood Pressure (ABP)	Invasive arterial line / NIBP	125–500 Hz	Systolic, diastolic, MAP, waveform morphology	Direct hemodynamic monitoring
Electrocardiogram (ECG)	Holter / bedside monitor	250–1000 Hz	RR interval, ST segments, HRV, QRS complex	Arrhythmia and ischemia detection
Pulse Oximetry (SpO2)	Photoplethysmography sensor	60–125 Hz	Oxygen saturation, pleth waveform	Oxygenation and perfusion status
Electroencephalography (EEG)	Depth-of-anesthesia monitor	256–512 Hz	Spectral indices, burst suppression, BIS	Anesthetic depth, cerebral monitoring
Electronic Health Records (EHR)	Hospital information system	Episodic	Comorbidities, medications, lab values, ASA class	Patient risk stratification
Anesthesia Records	Anesthesia information mgmt sys.	1–60 min intervals	Drug doses, airway events, fluid balance	Intraoperative management context
Capnography (EtCO2)	Sidestream/mainstream monitor	25–100 Hz	End-tidal CO2, respiratory rate, waveform	Ventilation and metabolism tracking
Invasive Hemodynamic (PA/CVP)	Pulmonary artery / CVP catheter	125–250 Hz	Cardiac output, SVR, PCWP, CVP	Advanced hemodynamic profiling

3.1. Continuous Physiological Signal Modalities

Continuous physiological signals measured throughout an operative procedure can provide the most frequent measurements (temporal resolution) available for predicting what will happen later during surgery (intraoperatively). One of the best examples of this is with the use of invasively measured arterial blood pressure (ABP) waveforms which are normally sampled between 125 Hz to 500 Hz. These measurements provide beat-to-beat measurements of hemodynamics (how the heart and blood vessels work) including systolic, diastolic, and mean arterial pressure (MAP). They also provide measures of ABP waveforms such as the area under the curve (AUC), pulse pressure (PP), and characteristics associated with the dicrotic notch. All of these measures can be used to describe how well blood is being pumped through your body based on vascular resistance, cardiac strength, and volume status [31]. An example of this is Edwards Lifesciences' Hypotension Prediction Index (HPI) algorithm which uses primarily continuous ABP waveform characteristics for developing its predictive models. In their prospective clinical trials, they reported predictive accuracy (defined as area under the receiver operating curve or AUROC value) greater than 0.93.

The continuous 12-lead ECG and/or continuous single-lead ECG provide complementary data on cardiac rhythm and conduction, as well as evidence for ischemia. The heart rate variability (HRV) metrics that can be calculated using the RR interval series derived from the ECG include timeliness domain metrics such as the standard deviation of the normal RR intervals (SDNN) and root mean square of the differences between adjacent normal RR intervals (RMSSD) and the frequency domain metrics of low (LF) and high (HF) frequency power, each of which provide a noninvasive view into the autonomic nervous system tone, which is an important factor in determining hemodynamic stability while under anaesthesia [32]. SpO2 pulse oximetry plethysmography waveforms provide peripheral perfusion and vascular tone information, but have considerably lower specificity for clinical diagnosis. The depth-of-anesthesia index values derived from EEG recordings, such as the Bispectral Index (BIS) and entropy index values, provide information about the cerebral state and depth of anaesthesia, both of which have an important influence on the human body's hemodynamic response to surgical stimulation.

3.2. Electronic Health Records and Patient History Data

EHR information is vital to patient care, but it is also a distinctive type of data. EHRs have a lot of meaning (or semantic richness), sampling is not happening at regular time periods (or irregular temporal sampling) and there is a high degree of variation between institutions' EHR data (or inter-institutional heterogeneity) [33]. EHRs have key variables that help predict patients' risk for developing IHI: (1) comorbidities (like heart failure, hypertension and/or diabetes) present before surgery; (2) patients' hemodynamic status (baseline; i.e. before surgery); (3) medications taken before surgery (of whatever type, including particularly antihypertensives, beta blockers and ACE inhibitor medications); (4) laboratory tests (hemoglobin, creatinine, electrolytes and coagulation tests); (5) history of previous surgical procedures or anesthetic administration; and (6) ASA physical status classification and/or procedure risk codes from EHR data. The ASA physical status classification system and procedure risk classifications provide rough risk stratification input (i.e., very accessible information on how patients were rated before surgery and how likely they are to have some type of risk following surgery, either from surgical or anaesthetic processes).

To combine EHR data with time-varying intraoperative signals or data, advanced preprocessing pipelines must be created to combine episodically sampled structured data with simultaneously collected high-frequency time-series data;

specifically, how the structured categorical data are converted into free text form label and vice-versa; and to address the high proportion of invalid data points (or missing data). NLP methods, such as ClinicalBERT and BioBERT, have been used extensively to extract structured clinical data from free-text anesthesia records or notes to increase the size of the multimodal data space for individualized patient risk prediction.

3.3. Anesthesia Information Management Systems

Anaesthesia Information Management Systems (AIMS) are a very specific type of clinical data system which gathers the types of detailed records needed to document the precise timing of any intraoperative medications (anesthetic agents), fluids or blood products administered to patients as well as detailed documentation from nursing staff and/or providers of care (35). The type of information captured in a standard AIMS database includes, but is not limited to, times from bolus doses of anesthetic agents and infusion rates; doses of neuromuscular blocking agents; doses of vasopressors and inotropes; and volumes of crystalloid and colloid fluids delivered; records of blood transfusions; and events related to airway management. Along with the pharmacological context of the patient, this information will assist in accurately interpreting the time-series of hemodynamic indices and thus assist in developing models which will differentiate the expected physiological responses to therapeutic interventions from spontaneous hemodynamic deterioration.

AIMS data, when combined with physiological signals, provide the basis for developing accurate, contextual prediction models. An example would be a prediction model for detecting a decrease in arterial pressure where the model includes both the administered propofol bolus and the concurrent arterial pressure waveform. The prediction model would then be able to distinguish between the decrease in arterial pressure as a result of normal pharmacodynamic response versus as a result of pathological deterioration, potentially reducing the number of false alarms due to prediction models incorrectly identifying patients at risk for an acute change in hemodynamic status producing a false alarm and improving clinical specificity of the prediction model (36).

3.4. Data Preprocessing, Synchronization, and Feature Extraction

The preprocessing of multimodal intraoperative data involves several technically complex stages that are critical in determining subsequent model performance [37]. Signal preprocessing of high-frequency waveforms consists of artifact detection/removal (motion artifacts, disconnections, electrosurgical interference), bandpass filtering, beat detection algorithms (for ABP and ECG) and resampling to standard frequencies. Temporal synchronization of heterogeneous data streams also requires precise timestamp alignment across monitoring systems with potentially different clocking and data export protocols—a challenge that is often underappreciated in retrospective data curation initiatives.

Generally, there are numerous different feature extraction strategies, including handcrafted clinical features (ie HRV metrics, pulse pressure variation, systolic pressure variation and pleth variability index) and learned representations using deep autoencoders/self-supervised training of unlabeled waveforms [38]. The sliding windows approach usually uses windows of 1-10 minutes, with steps of 10-30 seconds, in order to create prediction instances from continuous data streams with the labels based on the occurrence of hemodynamic instability events in a defined prediction horizon (typically 5-15 minutes ahead). The careful construction of these prediction windows (i.e., appropriate prediction horizons, minimum event durations, inclusion of surrounding intervention data) is a crucial methodological consideration that directly affects the clinical relevance and interpretability of model outcomes.

4. Multimodal Deep Learning Model Development and Architecture

A comparative overview of deep learning architectures relevant to intraoperative prediction is presented in Table 2.

Table 2: Comparison of major deep learning architectures, their primary mechanisms, optimal input types, relative strengths, and limitations in clinical perioperative applications.

Architecture	Primary Mechanism	Input Type	Strengths	Limitations
Convolutional Neural Network (CNN)	Spatial feature extraction via convolution filters	2D waveforms, images, spectrograms	Robust local pattern detection; computationally efficient	Limited temporal long-range dependency modeling
Recurrent Neural Network (RNN)	Sequential hidden state propagation	Time-series, sequential EHR	Captures temporal dependencies in sequential data	Vanishing gradient; slow training on long sequences
Long Short-Term Memory (LSTM)	Gated memory cells retaining long-term context	Time-series, longitudinal records	Excellent for long-horizon temporal modeling	High parameter count; memory-intensive
Transformer (Self-Attention)	Multi-head attention across full sequence	Multimodal tokens, EHR, signals	Global context modeling; parallelizable training	Data-hungry; computationally expensive
Graph Neural Network (GNN)	Node-edge message passing on graph structures	Clinical knowledge graphs, EHR relationships	Models relational structure between clinical variables	Graph construction non-trivial; interpretability limited
Hybrid CNN-LSTM	CNN for spatial + LSTM for temporal modeling	Waveform sequences, multimodal time-series	Combines local and sequential feature learning	Increased architectural complexity
Multimodal Transformer	Cross-modal attention alignment	Heterogeneous clinical modalities	Flexible fusion of disparate data types	Requires large paired multimodal datasets

Dataset characteristics and preprocessing strategies for key studies in this field are summarized in Table 3.

Table 3: Descriptive characteristics of principal datasets employed in multimodal deep learning studies for intraoperative hemodynamic prediction, including data modalities, clinical settings, prediction targets, and preprocessing approaches.

Dataset / Study	N (Cases)	Data Modalities	Setting	Outcome Predicted	Preprocessing Approach
MIMIC-III / MIMIC-IV	~60,000 ICU stays	EHR, ABP, ECG, labs	ICU / perioperative	Hemodynamic instability, mortality	Normalization, imputation, windowing
VitalDB (Korea)	6,388 surgical cases	ABP, ECG, SpO2, EtCO2, EHR	Operating room (intraoperative)	Intraoperative hypotension	Signal resampling, artifact removal, z-score
PhysioNet Hemodynamic DB	~10,000 records	ABP waveform, HR, SpO2	ICU/OR	Hemodynamic collapse, hypotensive episodes	Bandpass filtering, beat detection, feature extraction
Institutional Anesthesia EHR (UCSF/Stanford)	15,000–80,000 cases	Anesthesia records, EHR, vitals	Tertiary surgical centers	Intraoperative adverse events	Structured extraction, missing data imputation
Hypotension Prediction Index (HPI) Trials	~1,000 ICU/OR patients	Arterial waveform (continuous)	Multi-site OR and ICU	Hypotension within 5–15 min window	Waveform morphology features, real-time streaming
eICU Collaborative Research DB	~200,000 ICU admissions	EHR, lab values, vitals, treatment records	Multi-site ICU (USA)	Organ failure, hemodynamic deterioration	Multi-site harmonization, temporal alignment

4.1. Design Principles for Multimodal Neural Networks

Multimodal neural networks (MNNs) are used to make predictions about a patient's current clinical condition. The architecture of an MNN should maintain the unique characteristics of its different data types but also take full advantage of the ways in which those different data types support one another [39]. One common way to accomplish this is through a single network that is made up of multiple modality-specific encoder networks that encode their specific data types using encoders that are optimized for the characteristics of each of the multiple data types. For example, continuous waveform data will use either one dimensional convolutional neural networks (CNNs) or a combination of CNN and long short-term memory (LSTM) networks. Sequentially collected EHR data would use an embedding encoder with positional encodings to represent relationships within time and space. The third type of data—static preoperative features—would use a dense feed-forward encoder network to create dense vector representations that can undergo additional processing before a prediction is produced via a classification or regression layer.

For the first type of data—continuous waveform data—one-dimensional CNNs will extract local morphological features of the continuous waveforms represented in the data. Following the extraction of these local features, the extracted local morphological features will be modelled as a temporal series of morphological features using LSTM networks across the entire prediction window. A fourth alternative will be to use a transformer encoder with clinical-domain specific positional encodings when modelling the temporal from multiple samples across very long prediction windows, due to the self-attention mechanism of transformer encoder networks, providing superior performance than the sequentially weighted hidden-state mechanisms of LSTMs [40]. For the second type of data—EHR data—multiple bits or one-hot encodings of diagnosis and medication codes will be used in combination with dense vectors created from the continuous embedding of vital signs and laboratory values, to provide a dense vector representation of EHR data that can be further processed using attention-based MNN fusion modules.

4.2. Feature Representation, Embedding, and Temporal Modeling

To accurately represent features in a multimodal deep learning application within a clinical setting an understanding of how to derive informative representations from continuous physiological signals is needed. Such an understanding can come from any of the aforementioned methods for representation learning described above, such as convolutional autoencoders, contrastive learning methods (for example, SimCLR modified for clinical time series data), and spectral-domain representations like short-time Fourier-transformed spectrograms and continuous wavelet-transformed signals; all of which have produced informative, compact representations of physiological signals that reflect both morphometry (i.e., shape of the physiological signal) and statistics (i.e., differential variability and trending), and, thus, are jointly predictive of impending hemodynamic instability.

To account for the sequential nature of how the physiological signals change over time before impending IHI, high-dimensional temporal modeling architectures will be required that can simultaneously account for multi-scale patterns produced at various temporal scales (e.g., changes between beats across a -0.5 sec time scale vs hemodynamic trends across a several minute time scale vs the long-term drift of the physiological signal over several minute time scale); various multi-scale temporal modeling approaches currently exist, including multi-resolution CNNs with respect to differing window lengths, hierarchical long short-term memory (LSTM) networks with respect to varying state sizes or temporal convolutional networks (TCNs) with respect to dilated convolutions.

4.3. Handling Missing and Noisy Data

There are significant challenges associated with missing and noisy data when training and deploying deep learning models in the perioperative environment [43]. Patient movement, electrosurgical interference, disconnection of wires, and failure of sensors all generate signal artifacts that unfortunately result in a set of non-reliable epochs. As a result, if not managed correctly, they will decrease the performance of the model significantly in the situation where they are included. A number of algorithms can detect noise segments such as rule-based artifact detectors and anomaly detection models that were developed through training, and these algorithms are used to help filter and/or impute the corrupted data before training the model; therefore, they include both pre-processing and processing stage components.

Missing EHR variables, which are common for retrospective datasets because of the variability in indications of clinical guidelines for obtaining tests, can be handled with a variety of strategies, including mean/median imputation, regression-based imputation, k-nearest neighbor imputation and deep learning-based imputation using generative adversarial networks (GANs) or variational autoencoders (VAEs) [44]. Models built for real-time applications incorporate specific accommodations for missing data such as masked self-attention transformers that allow for dynamic attention to available input tokens only and graph-structured networks that enable the propagation of data across the available modalities; therefore, this ensures successful implementation of inference in clinical practice under the incomplete data conditions that exist before data can be collected and processed.

4.4. Training, Validation, and Optimization Strategies

The training of multimodal deep learning models for IHI prediction requires careful monitoring of class balance, temporal data leakages, and optimization stability. Intraoperative hypotension (IHI) events typically occur in 10-25 percent of the total intraoperative duration in unselected surgical populations; therefore, class imbalance problems can lead to outcome prediction bias in favour of the major class of negative hypotension, which occurs 75-90 percent of the time [45]. Class imbalance fixes include oversampling of the minority classes (SMOTE), undersampling, using weighted classes (weighted binary cross-entropy or focal loss) and using thresholds at inference.

Careful attention is needed to avoid temporal data leakage (using future time points during training labels and predictors) Many prediction tasks are time-series and require rigorous temporal train-validation-test splits with splits done at the patient-level (rather than time-window splits) to avoid leakage of information between cases for validity [46]. Optimization is provided using adaptive gradient methods (Adam, AdamW), learning rate scheduling (cosine annealing, cyclical), and regularization (dropout, L2 weights, and augmentation - timescale, amplitude, and Gaussian noise) to improve generalization. Bayesian search (& population-based training) rather than grid search is being used for the large hyperparameter search space found in multimodal architectures.

5. Performance Evaluation and Model Validation

A comparative summary of model performance across key studies is provided in Table 4.

Table 4: Performance metrics of multimodal and comparative models across key intraoperative hemodynamic prediction studies, including AUROC, sensitivity, specificity, F1 score, positive predictive value, and clinical prediction lead time.

Study / Model	Architecture	AUROC	Sensitivity	Specificity	F1 Score	PPV	Lead Time
Hatib *et al.* (HPI)	Gradient Boosting + waveform features	0.95	88%	87%	0.87	86%	5–15 min
Kendale *et al.*	Deep neural network (EHR)	0.85	79%	82%	0.80	78%	Pre-induction
Wijnberge *et al.*	HPI Algorithm (multimodal)	0.93	86%	85%	0.85	84%	10 min
Moghadam *et al.*	CNN-LSTM hybrid (waveform + EHR)	0.91	84%	88%	0.86	85%	10 min
Lee *et al.* (VitalDB)	Transformer (multimodal)	0.92	85%	89%	0.87	86%	5 min
Logistic Regression (baseline)	Statistical model	0.72	67%	74%	0.70	68%	N/A
Random Forest (ML baseline)	Ensemble learning	0.79	73%	78%	0.75	72%	N/A

5.1. Evaluation Metrics for Clinical Prediction Models

To evaluate deep learning models for IHI prediction, choosing metrics that consider the clinical application of a model as well as its statistical application is critical [47]. AUROC is the most common discrimination metric reported for models in the literature - it provides an indicator of a model's ability to accurately classify all positive cases as higher (rather than lower) than any negative cases, without regard to where the model's determined cut-off point would fall. AUROCs that fall within the 0.90 range or greater are clinically relevant as it relates to IHI prediction, and the vast majority of the best-performing MDL models fall within the 0.91–0.95 range, which is considerably higher than the AUROC levels (0.72–0.79) exhibited by conventional statistical and machine-learning baseline models.

In terms of deployment contexts, sensitivity (true positive rate) and positive predictive value (PPV) are of particular importance, because they directly impact both how many true hemodynamic events are detected (clinical safety) and how many of the alerts are clinically actionable (alert specificity). Because false-positive alerts contribute to alarm fatigue, creating an independent safety risk, having high PPV without sacrificing sensitivity is a major design requirement [48]. The area under the precision-recall curve (AUPRC) is particularly informative in the context of class imbalance, since it provides information about the model’s ability to find events and the precision of the model over its full range of operational thresholds. Calibration metrics (e.g., the Brier score and reliability diagrams) tell us how closely the predicted probabilities correspond to the actual event rates, which is critical for models that are deployed as probability estimators rather than binary classifiers.

5.2. Cross-Validation and External Validation Approaches

The way IHI prediction models are vetted is an extremely important aspect for evaluating how reliable the reported performances are, as well as assessing the level of confidence that could be placed on whether they are ready for use clinically [49]. Internal cross-validation methods (i.e. k-fold and nested cross-validation) that are typically used in model development, including when tuning hyperparmeters jointly with estimating the true effect, will not yield any information about how a model would perform on patients from different institutions, geographical areas, or points in time. Since it has been shown that deep learning algorithms are sensitive to many factors that can affect an input data distribution (e.g., types of physiologic monitoring devices, patient demographic characteristics, institutional protocols, and AIMS data structures); performing an external validation study using prospectively collected, independently acquired datasets is required for any use of an algorithm in a patient care setting.

The leading validation frameworks that have been utilized to date are: (1) LOIO cross-validation, which tests the performance of the model on data from the institution that was left-out while training on data from all the other institutions; this provides a rigorous measure of cross-institutional generalizability of the developed model [50]; and (2) The VitalDB dataset, which has been extensively utilized as an independent validation database for models developed on the MIMIC or PhysioNet databases in the United States, revealed clinically meaningful reductions in performance in models due to differences in patient demographics, surgical procedure types, and how physiologic monitoring devices are configured across different institutions.

5.3. Comparison with Conventional Statistical and Machine Learning Models

Multimodal deep learning approaches have been shown by comparative analyses to be superior to both traditional statistical methods such as logistic regression and Cox proportional hazards models and classic machine learning approaches including random forest models, gradient boosting machines and support vector machines in their ability to predict intraoperative hypotension [51]. Traditional statistical models generally result in AUROC values varying between 0.70 - 0.78, primarily due to the inability of these models to properly model the nonlinear dynamics and high-order feature interactions that exist within intraoperative physiological data. While handcrafted features created with domain expertise provide gradient boosting approaches (e.g., XGBoost, LightGBM) with competitive performance (AUROC values of 0.79 - 0.85), these approaches are much more dependent on expensive domain expertise for feature engineering than are multimodal deep learning methods and leave much of the information value contained in raw waveform data on the table.

The improvements in predictive accuracy demonstrated by many studies using ablation methods have consistently shown that multimodal deep learning methods outperform single-modality deep learning methods. Specifically, when discrete electronic health record data collected prior to surgery is combined with waveform signals collected while the patient is in the operating room, the AUROC improves between 0.03 - 0.07 compared to unipolar waveform-only models; furthermore, the addition of anesthesia administration records results in additional improvements in the accuracy of predictions made more than five minutes before the onset of intraoperative hypotension [52]. The empirical AUROC improvements documented in multi-modality research further confirm the theoretical justification for multimodal data integration and substantiate the added complexity and data infrastructure required to construct multimodal deep learning systems.

5.4. Interpretability and Explainability of Deep Learning Models

A major barrier to clinical trust and regulatory approval for deep learning models in high-stakes perioperative settings is the black box nature of these models [53]. Explainability techniques have been employed to enhance the transparency and clinical interpretability of IHI prediction models, including gradient-based attribution techniques (e.g., GradCAM, Integrated Gradients, and SHAP for deep models) that assign predicted model values to specific input features or slices of time so that the physiological signals and slices of time that are the best predictors of hemodynamic events can be identified. SHAP values, in particular, are typically used to provide local accurate and consistent feature attributions and then displayed in clinical dashboards to communicate model reasoning to anesthesiologists.

For transformer-based models, attention visualization provides an intrinsic means of interpretability by reflecting high attention weights to reflect the slices of time or modality pairs of inputs that most heavily influenced a given prediction. However, it is also important to understand that high attention weights should not be viewed as formal causal explanations for an association between inputs and predicted outputs and therefore should be interpreted with appropriate thoughtfulness [54]. The development of inherently interpretable models, including prototype-based networks, attention-augmented decision trees, and monotonic neural networks constrained by clinical prior knowledge, is a developing and important research opportunity.

6. Clinical Integration and Real-Time Implementation

6.1. Technical Requirements for Real-Time Deployment

There are engineering challenges that need to be solved to translate a successfully developed, trained machine learning decision logic (MDL) model into a clinically usable, real-time prediction system [55]. Real-time inference needs to occur in clinically actionable latencies—an example of which would be less than one second for inference each prediction window—while not disrupting the normal operation of connected monitoring systems. This means that the inference model architectures (e.g., quantized, pruned and distilled networks), the inference engines (ONNX Runtime, TensorRT) and the supporting hardware have to be designed with efficiency, and be integrated and/or co-located with the monitoring station.

Data pipeline streaming must work reliably with the various outputs from heterogeneous bedside monitors; this means being able to cope with variable rates of incoming data, cope with missing packets, and cope with all interruptions to the system, through graceful degradation, rather than total failure of the system. The integration with the hospital AIMS system will require standardised data exchange protocols, and HL7 FHIR (Fast Healthcare Interoperability Resources), has become the accepted standard for real-time clinical data exchange in the perioperative environment. The interface that will provide the prediction output will need to be carefully designed to present not only the predicted probability, but also the lead time of that prediction in such a way that does not increase the final cognitive burden on the anaesthesiologist.

6.2. Integration with Operating Room Decision-Support Systems

For the effective integration of MDL-based prediction into the operating room workflow, there must be thoughtful co-design between the clinical end-user and the decision support so that the clinical end-user sees it as a tool that assists them in their clinical responsibility rather than as an intervening or authoritative system. The various implementations of predictive decision support fall into two categories; passive outputs, such as risk score displays on anesthesia workstations, monitoring screens, and active tiered alert systems that provide a progressive escalation from visual notification to auditory alarm based on the probability of prediction of hemodynamic instability and the trend of the prediction. The integration of predictive output with AIMS allows the documentation of the hemodynamic risk history to be part of the anesthesia record, which can be examined for quality review of postoperative outcomes.

The advanced decision-support systems include not only the prediction of hemodynamic instability but also propose model-based recommendations for therapeutic interventions, such as suggested dosing of vasopressors, volumes of fluid bolus, and anesthetic depth based on the predicted hemodynamic instability, that are generated by reinforcement learning agents trained on historical outcomes. These closed-loop decision-support systems will be the future of perioperative AI, but there needs to be a significant amount of clinical validation, regulatory oversight, and development of liability frameworks for these systems prior to deploying them in widespread use.

6.3. Human-AI Interaction and Clinician Trust

The level of trust that clinicians have in MDL-based decision support systems and how they use them will ultimately determine how successful they are in the operating room. Trust is determined by clinicians' ability to see how the predictions were made (prediction transparency), their confidence in the predictions (calibration confidence), whether or not there are clear pathways for responding to alerts (actionability), and how well the alerts are calibrated to actual events (historical accuracy), based on ongoing feedback about the outcomes of alerts. Alert fatigue and adoption barriers can frequently be attributed to excessive alert frequency, poor calibration, or the inability to explain predictions.

To ensure that MDLs are deployed responsibly, it is also essential to include human factors engineering techniques such as usability testing with anesthesia providers in realistic simulated OR environments, iterative interface design, and use implementation science frameworks to facilitate institutional adoption, as part of a responsible strategy for MDL deployment. The concurrent development of clinical decision algorithms, which specify the appropriate clinical response for different levels of predicted hemodynamic risk, should also be considered a critical component of the overall MDL development process.

6.4. Evidence from Clinical Implementation Trials

The best proof for how MDL-based hemodynamic prediction can affect patients in a clinical setting comes from randomized controlled trials that measure how effective MDL systems are when deployed in real time. The PREVENT study found that there was a statistically significant difference in the amount of time during the surgery that the MAP was below 65 mmHg in the experimental vs. control group. The Wijnberge study found that the HPI-guided protocol approach had about a 50% greater reduction in how often and for how long the patients experienced hypotensive events than the conventional system.

Table 5 shows data from clinical studies that implemented MDL prediction systems and demonstrate both that MDL prediction systems are capable of producing clinical outcomes and are complex with regard to how much of the positive outcome can be attributed to the predictive component vs. the protocolized intervention component of the MDL system. Future research using factorial trials, adaptive protocols, and long-term (30 day) mortality and adverse kidney injury (AKI) and myocardial injury after non-cardiac surgery (MINS) data will be needed to better define the extent of clinical outcome improvements that can be attributed to multimodal predictive hemodynamic management.

Table 5: Summary of clinical implementation studies of multimodal and AI-based hemodynamic prediction systems, including study settings, interventions, primary outcomes, results, and identified adoption challenges.

Implementation Study	Clinical Setting	Intervention	Primary Outcome	Result	Adoption Challenge
Wijnberge *et al*. RCT, 2020	General OR, Netherlands	HPI-guided vasopressor protocol	Time-weighted average MAP <65 mmHg	Reduced hypotension duration by 50%	Alert fatigue; workflow integration
Maheshwari *et al*. PREVENT RCT	Multi-center surgical OR	Real-time HPI alerts + intervention protocol	Intraoperative hypotension incidence	Significant reduction vs. standard care	Clinician trust, training burden
Edwards Acumen IQ Sensor Trial	Cardiac / high-risk surgery OR	Continuous SV/CO monitoring + prediction	Hemodynamic instability events, ICU LOS	Decreased vasopressor use, shorter ICU stay	Device cost, calibration requirements
EHR-based DL Prediction (Multi-site)	Large academic medical centers	Pre-operative risk score + intraop alert	Composite adverse hemodynamic events	Improved detection; no mortality benefit shown	EHR heterogeneity across sites
AI-Guided Anesthesia Pilot (Korea)	Tertiary university hospital OR	Multimodal DL decision support + nursing alert	Incidence of MAP <65 >5 min	Reduction in event severity and duration	Language/cultural workflow adaptation

7. Challenges, Limitations, and Ethical Considerations

7.1. Data Heterogeneity and Quality Issues

The current practical utilization of MDL models for IHI prediction faces a number of significant obstacles associated with the heterogeneity and variability in the quality of clinical data among various monitoring systems and between healthcare institutions. Data generated by bedside monitoring devices from multiple manufacturers (e.g., Philips, GE Healthcare, Dräger, Mindray) are characterized by differing sampling rates, algorithms for signal processing, and formats for data output that yield systematic disparities in the distribution of features. These systematic discrepancies may have a serious negative effect on the performance of a model that has been trained on data from a single monitoring environment and used at another environment. Also, the variations associated with the different AIMS platforms (Epic Anesthesia, Merge Emed, Metavision) and their mechanisms for exporting data introduce an amount of heterogeneity into the availability and structure of the pharmacological and procedural data.

Challenges to data quality include the frequent occurrence of signal artifacts in the arterial pressure waveform, the inconsistent manner in which AIMS documents patient information, the systematic nature of missing data from retrospective EHR datasets, and the inability to implement standardized practices for annotating clinical outcomes across institutions. To address these multiple challenges, there will need to be a substantial investment in infrastructure for curating data, establishing inter-institutional standards for the harmonization of data, and creating protocols for an effective prospective data collection strategy, designed to meet the training needs of MDLs. An important emerging technical priority is the development of a federated pre-processing pipeline that standardizes data quality control from multiple institutions, while avoiding the need for centralized data aggregation.

7.2. Model Generalizability and External Validity

Clinical MDL systems still face an ongoing challenge to model generalization for unique patient populations, surgical specialties, geographic regions, and temporal timeframes. The reduction of performance on external validation cohorts compared to internal development data is a widely established finding in the literature. The average reduction in area under the receiver operating characteristic (AUROC) between evaluation of models built with data from one institution and externally to that institution is estimated to be 0.05-0.15. The major contributors to this performance reductions are differences in demographic characteristics of patients (i.e., age, comorbidity burden, body habitus), surgical case mix, anesthesia practice patterns, fluid management protocols, and the epidemiology of hemodynamic instability in each organization.

Domain adaptation techniques (i.e., fine-tuning pre-trained models with small amounts of data from the target institution), multi-site training with institution-level batch normalization or modulation layers, and ensemble approaches to combine predictions from multiple institution-specific models are all strategies to improve generalizability of clinical MDL systems. In particular, the use of transfer learning, in which there is fine-tuning of general pre-trained models on smaller, institution-specific datasets, has great promise for providing robust cross-institutional performance with minimal data and computational requirements for local adaptation.

7.3. Bias, Fairness, and Transparency

There are risks when using deep learning models in high-stakes clinical settings that the models will continue or amplify biases that were present in the historical training data used to build them. When using historical datasets that contain systemic disparities in monitoring, aggressiveness of treatment, or documentation of outcomes across demographic groups (e.g., race, ethnicity, gender, age, and SE status), models built on these datasets will generate predictions that are biased against populations who are not currently well-served. Algorithmic fairness, which involves assessing the performance of models separately for different demographic groups, is an important part of the process for validating models used in the perioperative literature on AI, but often is not included with the other model validation steps.

Transparency in model development and reporting is critical for allowing for a third-party's independent evaluation of the work, along with regulatory review, and for clinicians to make informed decisions about adoption of the model. Transparency includes providing as much detail on the characteristics of the training data used, the architecture of the model, hyperparameters, validation methods, and performance assessed across all relevant subpopulations. An excellent way to achieve this transparency is through adherence to widely accepted standards for reporting what is done, such as the TRIPOD+AI framework for reporting the use of multivariable prediction models and CONSORT-AI framework for randomised trials of AI interventions; at a minimum, authors should adhere to these reporting standards when publishing papers in this area.

7.4. Regulatory and Data Privacy Considerations

The Food and Drug Administration (FDA) regulates MDL-based clinical decision support tools used in operating rooms in the United States as Software as a Medical Device (SaMD). The FDA does this through the De Novo, 510(k), or premarket approval (PMA) pathways, depending on the risk classification of the device. The FDA's proposed rules for AI/ML-based SaMD include specific requirements for predetermined change control plans (PCCPs) that take into account the ability of adaptive AI systems to learn and update models over time. This is because traditional one-time device clearance is not enough for algorithms that are meant to change with new data. The Health Insurance Portability and Accountability Act (HIPAA) in the US, the General Data Protection Regulation (GDPR) in the EU, and

The Health Insurance Portability and Accountability Act (HIPAA) in the US, the General Data Protection Regulation (GDPR) in the European Union, and similar national laws all make it very hard to collect, store, send, and use patient-level physiological and clinical data for AI model training. Federated learning architectures, which enable model training across distributed institutional datasets without centralizing patient-level data, represent a technically mature and increasingly adopted approach to enabling large-scale multimodal model development while maintaining compliance with data privacy regulations.

8. Future Directions and Innovations

8.1. Personalized and Adaptive Prediction Systems

The next step in MDL-based IHI prediction is to move from static models that work for everyone to truly personalized, adaptive prediction systems that keep getting better at making predictions based on how a single patient is doing. Personalized models could include baseline physiological characteristics unique to the patient, historical response patterns to anesthetic agents, and probabilistic models of hemodynamic state that are updated in real time to reflect changes in the surgical context. Bayesian online learning and continual learning frameworks are promising technical ways to make this vision a reality. They let model parameters be updated little by little as new patient data comes in during a procedure.

Meta-learning approaches (learning to learn), wherein a model is trained to swiftly adapt to new patients or contexts utilizing a minimal amount of patient-specific data, are particularly pertinent in the intraoperative setting, where the total data available for personalization is inherently constrained. These methods use the preoperative baseline as a starting point and change predictions as early intraoperative physiological data comes in. This makes it possible to move smoothly from preoperative risk stratification to real-time intraoperative prediction within a single modeling framework.

8.2. Federated Learning and Privacy-Preserving Data-Sharing Frameworks

Federated learning (FL) is a new way to train MDL models on large, varied clinical datasets while still protecting patient privacy and following institutional data governance rules. In FL, each participating institution trains its own model on its own patient data. Only model parameter updates (gradients) are sent to a central aggregation server, not raw patient data. The aggregated model takes advantage of the variety of the federated training data without having to send any individual patient record outside of the institution that created it. This meets both privacy and institutional data sovereignty standards.

Recent implementations of federated learning in perioperative AI tasks have shown that FL models can attain performance comparable to centrally trained models, especially when aggregation strategies (FedAvg, FedProx, FedBN) are employed to address the non-IID (non-independently and identically distributed) characteristics of clinical data across institutions. Communication efficiency, differential privacy guarantees, model poisoning resistance, and client heterogeneity management are still being researched. The Federated Tumor Segmentation (FeTS) initiative and the HealthChain FL consortium are two new infrastructure models for large-scale federated clinical AI development that could be used for perioperative prediction applications.

8.3. Wearable and IoT-Based Perioperative Monitoring Integration

Utilizing wearables and IoT-based monitoring devices for perioperative predictions offers a means to expand hemodynamic surveillance into the preoperative and postoperative phases, thereby allowing for longitudinal modeling of patients' physiological trajectory throughout the entire perioperative continuum. Wearable continuous devices that measure photoplethysmography, accelerometry, skin temperature, galvanic skin response, and 1-lead ECG will create a robust preoperative physiological characterization that can dramatically increase prediction accuracy and personalization, when combined with intraoperative monitoring data.

Integration of the Internet of Things (IoT) into clinical settings presents significant technical obstacles, which have become easier to overcome as wearable sensor technology has developed. Some of these obstacles include interference of wireless signals and electromagnetic fields in operating rooms, limitations on battery life, device calibration and standardization, and the need for interoperability with existing hospital information systems. Edge computing technologies that do local signal processing and feature extraction on wearables prior to sending compressed feature vectors to the central prediction system are addressing interoperability and privacy constraints in the integration of IoT devices into clinical settings.

8.4. Explainable AI and the Future of Clinical Decision-Making

New explainable artificial intelligence (XAI) technologies that offer useful, clinically relevant, action-oriented explanations of MDL predictions possibly represent the most significant technical challenge for safe, trustworthy adoption of these new technologies in the clinical setting. Current XAI methodologies (post-hoc attribution algorithms such as SHAP, LIME; attention visualizations) provide approximate explanations to enable auditing and educating end users (the model users) about how the model is producing its results; however, the above-mentioned methods do not provide the causal, mechanistic explanations that a physician will need to accept the use of AI-generated alerts for clinical decision making in high-stakes situations.

New techniques such as causal machine learning systems (causal representation learning and counterfactual explainability), physics-informed neural networks which include physiological mathematical models as structural constraints, and inherently interpretable neuro-symbolic systems that integrate neural feature extraction with symbolic reasoning have the potential to create AI systems whose reasoning aligns with clinical physiology as being genuine. The final aim of an AI system will be to predict when an individual will experience hemodynamic instability while being able to articulate in clinically relevant terms the specific physiological and pharmacological variables that are contributing to the heightened risk of that patient at that moment; thus turning the AI from an opaque oracle into a transparent partner in clinical decision making.

9. Conceptual Diagrams and System Architecture Illustrations

The following figures illustrate key conceptual frameworks discussed in this review. All figures are schematic representations designed to convey architectural and workflow concepts; they do not represent graphical data visualizations.

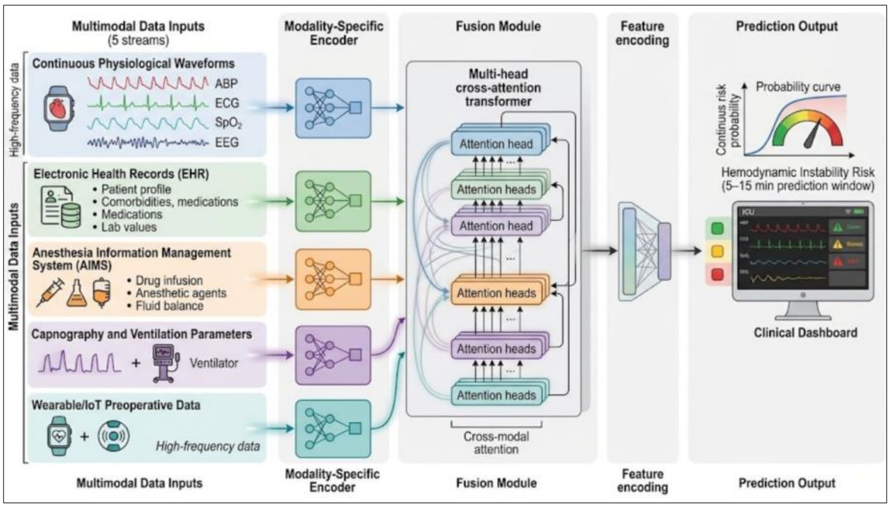

Fig 1: Conceptual Diagram of Multimodal Data Integration for Intraoperative Prediction

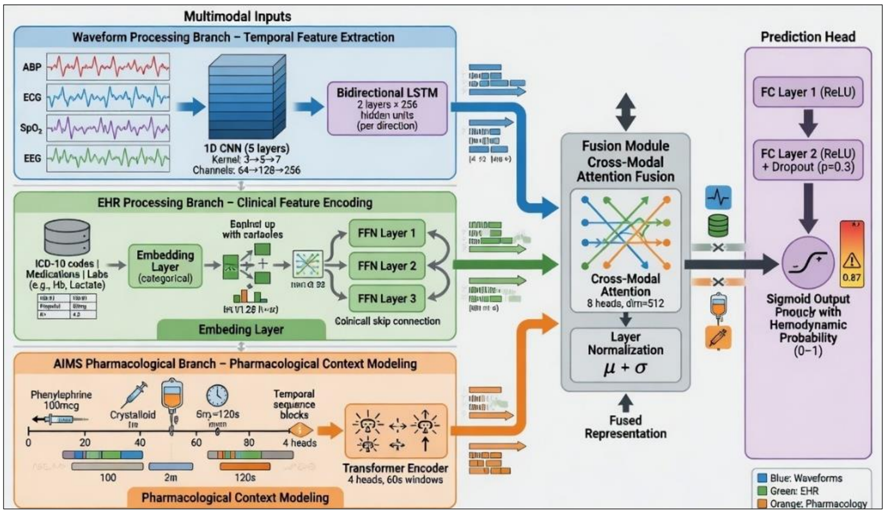

Fig 2: Architecture of the Proposed Multimodal Deep Learning Model

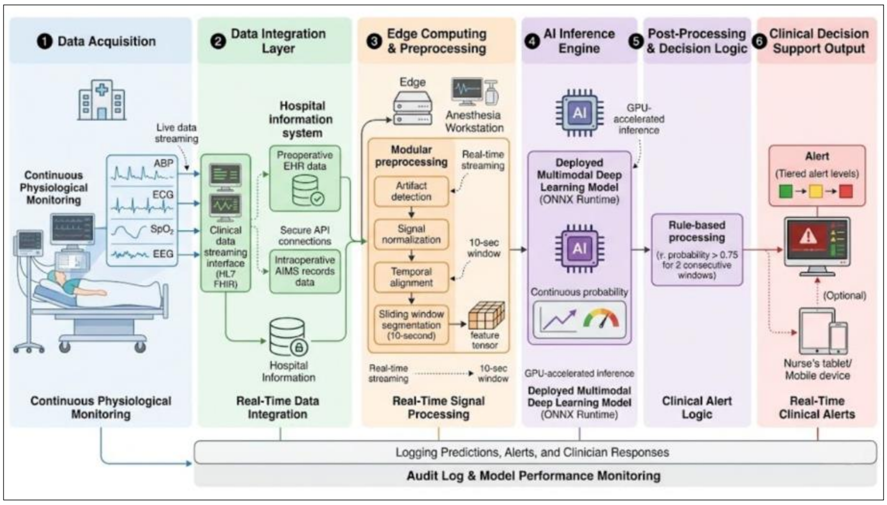

Fig 3: Workflow of the Intraoperative Hemodynamic Prediction System

10. Conclusion

This thorough analysis gathered the current data about the state of multimodal deep learning for predicting hemodynamic instability during surgery. The comprehensive review included: the epidemiology of perioperative hemodynamic instability, the theoretical basis for multimodal deep learning, all available types of data and how to preprocess those data, advances in model architecture and train techniques, the evidence for clinical performance and the implementation of multimodal deep learning, and challenges and ethical issues regarding using artificial intelligence solutions in high-risk surgical environments, all providing a comprehensive overview of this field.

The conclusion is that multimodal deep learning models which utilize continuous physiological waveforms with electronic health record data, including the context of the administration of anesthesia, demonstrate better prediction performance (0.91-0.95 AUROC), as well as clinically significant lead time (5-15 minutes) compared with traditional models. Prospective randomized trials support that using multimodal deep learning models to assist in predicting hemodynamic instability when combined with pre-planned pathways for the delivery of appropriate interventions can reduce both the incidence and length of intraoperative hypotension by as much as 50% (and), as a result, may reduce the risk of organ dysfunction such as injury to the heart, acute renal failure (AKI) and postoperative cognitive dysfunction.

Recommendations for the implementation of clinical practices involve using standardized methods to collect multimodal data in the operating room, investing in HL7 FHIR compatible data streaming technology, validating the algorithms being deployed on diverse populations of patients before they are deployed, co-designing the alerts and responses with the anaesthetic provider, continually monitoring algorithmic performance while providing regular feedback loops to improve those algorithms, and evaluating the fairness of the algorithms prior to deployment using demographic disaggregation.

There is a significant gap in research that needs to be filled before MDLs can reach their potential for providing support during the perioperative period; specifically, multi-institutional, multi-site studies designed to capture data on thousands of patients and follow them for extended periods of time; the creation and standardization of a federated learning framework for cross-institutional model training (while protecting the privacy of each institution's data); evaluation of new personalized adaptations to current predictions for each patient during the surgery; and developing methods to make the results of AI algorithms more understandable and usable by the clinical provider.

Long story short, adding multiple modes together for deep learning purposes in monitoring during an operation brings a new phase into anesthesia, changing it from a discipline that reacts when something goes wrong, to one that predicts and prevents problems before they happen. Making this all happen requires advanced machine learning techniques along with clinical testing, engineering of work done by people, ethics and how people will use AI to enhance but not replace the expert clinician's decisions.

References

1. Sessler DI, Bloomstone JA, Aronson S, *et al*. Perioperative Quality Initiative consensus statement on intraoperative blood pressure, risk and outcomes for elective surgery. Br J Anaesth. 2019;122(5):563-574.
2. Bijker JB, van Klei WA, Kappen TH, *et al*. Incidence of intraoperative hypotension as a function of the chosen definition: literature definitions applied to a retrospective cohort using automated data collection. Anesthesiology. 2007;107(2):213-220.
3. Wesselink EM, Kappen TH, Torn HM, *et al*. Intraoperative hypotension and the risk of postoperative adverse outcomes: a systematic review. Br J Anaesth. 2018;121(4):706-721.
4. Maas EHA, Bauer M, Tonner PH. Intraoperative hypertension. Best Pract Res Clin Anaesthesiol. 2021;35(2):209-217.
5. Aronson S, Dyke CM, Stierer KA, *et al*. The ECLIPSE trials: comparative studies of clevidipine to nitroglycerin, sodium nitroprusside, and nicardipine for acute hypertension treatment in cardiac surgery patients. Anesth Analg. 2008;107(4):1110-1121.
6. Joosten A, Desebbe O, Suehiro K, *et al*. Accuracy and precision of non-invasive cardiac output monitoring devices in perioperative medicine: a systematic review and meta-analysis. Br J Anaesth. 2017;118(3):298-310.
7. Walsh M, Devereaux PJ, Garg AX, *et al*. Relationship between intraoperative mean arterial pressure and clinical outcomes after noncardiac surgery: toward an empirical definition of hypotension. Anesthesiology. 2013;119(3):507-515.
8. Salmasi V, Maheshwari K, Yang D, *et al*. Relationship between intraoperative hypotension, defined by either reduction from baseline or absolute thresholds, and acute kidney and myocardial injury after noncardiac surgery. Anesthesiology. 2017;126(1):47-65.
9. Sun LY, Wijeysundera DN, Tait GA, Beattie WS. Association of intraoperative hypotension with acute kidney injury after elective noncardiac surgery. Anesthesiology. 2015;123(3):515-523.
10. Vernooij LM, van Klei WA, Machina M, *et al*. Different methods of modelling intraoperative hypotension and their association with postoperative complications in patients undergoing non-cardiac surgery. Br J Anaesth. 2018;120(5):1046-1056.
11. Varon J, Marik PE. Perioperative hypertension management. Vasc Health Risk Manag. 2008;4(3):615-627.
12. American Society of Anesthesiologists. Standards for Basic Anesthetic Monitoring. ASA Standards and Guidelines; 2020. Available from: https://www.asahq.org/standards-and-guidelines/standards-for-basic-anesthetic-monitoring.
13. Greenberg SB, Murphy GS, Vender JS. Current use of oxygen monitoring beyond the operating room. Curr Opin Anaesthesiol. 2011;24(6):649-654.
14. Lee TH, Marcantonio ER, Mangione CM, *et al*. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. Circulation. 1999;100(10):1043-1049.
15. Kendale S, Kulkarni P, Rosenberg AD, Wang J. Supervised machine-learning predictive analytics for prediction of postinduction hypotension. Anesthesiology. 2018;129(4):675-688.
16. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-444.
17. Hatib F, Jian Z, Buddi S, *et al*. Machine-learning algorithm to predict hypotension based on high-fidelity arterial pressure waveform analysis. Anesthesiology. 2018;129(4):663-674.
18. Davies SJ, Vistisen ST, Jian Z, *et al*. Ability of an arterial waveform analysis-derived hypotension prediction index to predict future intraoperative hypotensive events in patients having major abdominal surgery. Anesth Analg. 2020;130(2):352-359.
19. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, MA: MIT Press; 2016.
20. Rajpurkar P, Chen E, Faizi O, *et al*. AI in health and medicine. Nat Med. 2022;28(1):31-38.
21. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735-1780.
22. Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all

you need. Adv Neural Inf Process Syst. 2017;30:5998-6008.
23. Ngiam J, Khosla A, Kim M, *et al*. Multimodal deep learning. Proc Int Conf Mach Learn. 2011:689-696.
24. Aczon M, Ledbetter D, Ho L, *et al*. Dynamic mortality risk predictions in pediatric critical care using recurrent neural networks. arXiv:1701.06675. 2017.
25. Pereira RD, Hicks H, Joosten A, *et al*. Intraoperative fluid management: clinical impact and validity assessment. J Clin Anesth. 2020;65:109869.
26. Baltrusaitis T, Ahuja C, Morency LP. Multimodal machine learning: a survey and taxonomy. IEEE Trans Pattern Anal Mach Intell. 2019;41(2):423-443.
27. Ramachandram D, Taylor GW. Deep multimodal learning: a survey on recent advances and trends. IEEE Signal Process Mag. 2017;34(6):96-108.
28. Arevalo J, Solorio T, Montes-y-Gomez M, *et al*. Gated multimodal units for information fusion. arXiv:1702.01992. 2017.
29. Chen RJ, Lu MY, Wang J, *et al*. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE Trans Med Imaging. 2022;41(4):757-770.
30. Maheshwari K, Buddi S, Jian Z, *et al*. Relationship between hypotension prediction index and intraoperative hypotension during noncardiac surgeries. J Clin Monit Comput. 2021;35(1):65-72.
31. Meidert AS, Saugel B. Techniques for non-invasive monitoring of arterial blood pressure. Front Med (Lausanne). 2018;4:231.
32. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Circulation. 1996;93(5):1043-1065.
33. Rajkomar A, Oren E, Chen K, *et al*. Scalable and accurate deep learning with electronic health records. NPJ Digit Med. 2018;1:18.
34. Szolovits P, editor. Artificial Intelligence in Medicine. AAAS Symposium. Boulder, CO: Westview Press; 1982.
35. Springman SR, Bharagava S, editors. Anesthesia Information Management Systems. New York: Springer; 2015.
36. Sessler DI. Perioperative thermoregulation and heat balance. Lancet. 2016;387(10038):2655-2664.
37. Johnson AEW, Pollard TJ, Shen L, *et al*. MIMIC-III, a freely accessible critical care database. Sci Data. 2016;3:160035.
38. Lee HC, Jung CW. VitalDB: a high-fidelity multi-parameter vital signs database in surgical patients. Sci Data. 2022;9:279.
39. Esteva A, Robicquet A, Ramsundar B, *et al*. A guide to deep learning in healthcare. Nat Med. 2019;25(1):24-29.
40. Huang G, Tan JY, Chen MY, Elgendi M. Transformers for electrocardiogram classification. Front Cardiovasc Med. 2022;9:915689.
41. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE Conf Comput Vis Pattern Recognit. 2016:770-778.
42. Bai S, Kolter JZ, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv:1803.01271. 2018.
43. Ghassemi M, Naumann T, Schulam P, *et al*. A review of challenges and opportunities in machine learning for health. AMIA Jt Summits Transl Sci Proc. 2020;2020:191-200.
44. Yoon J, Jordon J, van der Schaar M. GAIN: missing data imputation using generative adversarial nets. Proc Int Conf Mach Learn. 2018;80:5689-5698.
45. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321-357.
46. Kaufman S, Rosset S, Perlich C, Stitelman O. Leakage in data mining: formulation, detection, and avoidance. ACM Trans Knowl Discov Data. 2012;6(4):1-21.
47. Steyerberg EW, Vickers AJ, Cook NR, *et al*. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21(1):128-138.
48. Schmid F, Goepfert MS, Kuhnt D, *et al*. The wolf is crying in the operating room: patient monitor and anesthesia workstation alarming patterns during cardiac surgery. Anesth Analg. 2011;112(1):78-83.
49. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.
50. Nestor B, McDermott MBA, Boag W, *et al*. Feature robustness in non-stationary health records: caveats to deployable model performance in common clinical machine learning tasks. Proc Mach Learn Res. 2019;106:381-405.
51. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B Methodol. 1996;58(1):267-288.
52. Moghadam MC, Sanchez-Londono M, Cannesson M, *et al*. Intraoperative hypotension prediction: a model comparison using a multi-modal deep learning framework. J Clin Monit Comput. 2023;37(1):101-114.
53. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1(5):206-215.
54. Jain S, Wallace BC. Attention is not explanation. arXiv:1902.10186. 2019.
55. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56.

How to Cite This Article

Vaibhav V, Sinha N, Vaid N, Maheshwari M, Madhrey M, Patel SA. Multimodal deep learning for intraoperative hemodynamic instability prediction. Int J Med All Body Health Res. 2026;7(2):45–58. doi:10.54660/IJMBHR.2026.7.2.45-58