Drug Overdose Vital-Signs Evaluator Using Machine Learning

Anush Niranjan Lingamoorthy[1] , **Abhishek Kumar Mishra**[1] , **Suman Kumar**[1] , **David Gordon**[2] , **Jacob Brenner**[3] , **Nagarajan Kandasamy**[1] , **Amanda Watson**[4]
[1]Drexel University
[2]Thomas Jefferson University
[3]University of Pennsylvania
[4]University of Virginia
{aln57, am4862, sk4266, nk78}@drexel.edu, jacob.brenner@pennmedicine.upenn.edu,
David.Gordon@students.jefferson.edu, aawatson@virginia.edu

Abstract

Opioid overdose is an escalating global epidemic, affecting 16 million individuals. Lack of overdose detection and slower response times are the leading causes of overdose deaths. During a fatal opioid overdose, the user exhibits motionlessness, lack of breathing, and hypoxemia (oxygen saturation drops). In this paper, we discuss the development of a shoulder-based wearable overdose detection device that monitors hypoxemia, motion, and respiration. The device's design considers the underserved socio-economic population and their psychological contexts. However, conventional approaches to detecting an overdose typically focus on a single biomarker. To address this, we have developed a robust capsule networks based machine learning (ML) model, OxyCaps that integrates oxygen saturation, respiration rate, and motion to classify different levels of hypoxemia. This also helps improve patient adherence by decreasing the chances of false positive alerts. To determine a hypoxemic state, the model considers various features like skin tone, body physiology, motion, and photoplethysmography (PPG) signals. In the absence of real-world opioid overdose data, our research leverages data collected by our device from 19 patients experiencing sleep apnea, exploiting the parallels between overdose and apnea biomarkers. Our dataset provides a novel compilation of raw PPG and motion signals detected from the shoulder. Our model classifies 3 stages of hypoxemia with an average accuracy of 92%, specifically achieving a high recall of 0.98 for the critical hypoxemic state that is crucial in determining an overdose.

1 Introduction

The devastating impacts of the global opioid epidemic resonate across societies and public health systems, with the annual death toll due to opioid overdose (OD) reaching a staggering 120,000 worldwide. An estimated 16 million [Chang *et al.*, 2018] suffer from opioid use disorder (OUD), highlighting the urgent need for innovative and effective interventions to curb this crisis. However, the nature of fatal overdose events poses substantial challenges to effective data collection, compounding the difficulties in understanding, predicting, and preventing these tragedies. These obstacles become more pronounced given that more than half of opioid users are known to consume opioids when alone [Ogeil *et al.*, 2018], further limiting the chances of timely detection and intervention during an overdose. Despite these grim realities, there is a silver lining: research indicates that 76% of those suffering from opioid use disorder would willingly wear a device on the shoulder capable of detecting an overdose event and alerting emergency responders [Kanter *et al.*, 2021a], provided it is easily concealable. Guided by this finding, our research explores developing and deploying a shoulder-based wearable device, addressing the opioid epidemic from a unique technological and social impact perspective.

The primary focus of this paper is to introduce a novel capsule network [Sabour *et al.*, 2017] OxyCaps architecture designed to classify hypoxemia during an overdose modeled using sleep apnea data. The severity of hypoxemia is classified into 3 levels based on SpO_2 – Normal (96%-100%), Moderate (92%-95%), and Severe (91%-88%). Clinically, a healthy individual or COPD (Chronic Obstructive Pulmonary Disease) patient with SpO_2 levels less than 92% and 88% respectively are considered hypoxemic [Lee *et al.*, 2000]. During an opioid-OD oxygen saturation (SpO_2), motion and respiration rate are affected due to opioid-induced respiratory depression [Boyer, 2012a]. These biomarkers are used as overlying features to determine the level of hypoxemia experienced by the patient. Our proposed device takes the form of a sensor that adheres to the deltoid, offering a discreet yet effective monitoring solution according to our opioid population's needs. Leveraging the capabilities of optical and motion sensors, this device generates photoplethysmography (PPG) and 6-axis motion signals, which can be processed to estimate the user's oxygen saturation, respiration rate, and motion levels. These biomarkers, coupled with additional generated features, form the foundation for our overdose detection system using hypoxemia as our indicator.

Capsule networks are adept at discerning spatial relationships among features and understanding temporal hierar-

chies, making them particularly effective at identifying patterns spanning various time scales. Such temporal intricacies might arise from filtering processes inherent to standard pulse oximeters used in collecting labeled data or from physiological lags across different sensing areas (finger and shoulder). This also enables OxyCaps to better handle temporal delays between the onset of respiratory depression and drops in SpO_2 during apneic events. To train the OxyCaps model for precise classification of hypoxemia indicative of overdoses, we sourced our dataset from patients diagnosed with sleep apnea using custom hardware. One of the biggest challenges in opioid-OD research is the scarcity of clean data pertaining to actual overdose events [Campbell *et al.*, 2023]. Instead of relying on hard-to-obtain data from fatal overdose incidents, we have chosen to build our dataset from patients suffering from obstructive sleep apnea (OSA). Symptoms of OSA — hypoxemia, decreased breathing, and motionlessness during an apneic event — bear significant resemblance to those of an opioid-OD [Boyer, 2012b], making OSA patient data an appropriate surrogate for modeling purposes. The presence of reactive body movements post-apneic events in OSA patients creates a prominent differentiator between OSA and opioid-OD-related hypoxemic events. During an opioid-OD, the individual is completely motionless, and their hypoxemic state tends to increase if the reversal agent naloxone is not administered immediately. Whereas, a patient with OSA will recover from their apneic event during which a resumption of breathing is exhibited and their hypoxemic state normalizes. This differentiator helps us use OSA patients as a reasonable model to develop an overdose classifier.

There is a critical need for a device that can not only detect opioid-ODs but also one that patients are willing to wear. To foster greater acceptance and utilization, we integrated a patient-centric iterative approach into our design process based on prior research of this underserved community [Kanter *et al.*, 2021b]. Through a combination of interviews, informal focus groups, and surveys, we addressed the unique needs and concerns of the OUD community. The result was a modular, easy-to-wear, and discreet shoulder-based sensor whose comfort and accuracy levels we validated using sleep apnea patients. The ability to detect overdoses via shoulder monitoring will also enable seamless integration with autoinjector technologies in the future that administer naloxone to reverse fatal overdoses [Chan *et al.*, 2021; Imtiaz *et al.*, 2021; Dhowan *et al.*, 2019; Lingamoorthy *et al.*, 2023]. At the end of the study, the patients were surveyed to better understand the wearability and comfort of our shoulder device during sleep compared to commercial pulse oximeters and OD monitors. The shoulder device was ranked higher than the commercial device in terms of comfort and wearability due to its small size, location, and application using an adhesive.

Our contributions are summarized as follows:

1. Development of a novel capsule network implementation OxyCaps with routing to capture spatial information between features for estimating hypoxemia.
2. Evaluation of the algorithm using over 150 hours of PPG, pulse oximetry, and motion data collected from 19 sleep apnea patients to simulate overdose events caused by sleep apnea episodes.
3. Provision of the uniquely acquired shoulder-based PPG and motion signal dataset from sleep apnea patients for public access [Kumar, 2024].

As our understanding of the opioid epidemic continues to evolve, the role of artificial intelligence in combating this crisis becomes increasingly evident. Our research represents a step forward in this domain, harnessing AI to improve data collection methods, develop nuanced problem models, and facilitate real-world testing and evaluation. By focusing on these key areas, we aim to contribute meaningfully to ongoing efforts to mitigate the devastating social impacts of the opioid crisis. Our study underscores the promise of AI-based solutions in addressing public health challenges, using innovative technology to enhance the safety and well-being of individuals living with opioid use disorder. By introducing a wearable, patient-centric device capable of detecting imminent overdose events, we can help pave the way for future interventions and policies to combat the opioid crisis more effectively. Our findings also carry broader implications for the field of medical technology and AI, highlighting the potential of these tools in creating more responsive, inclusive, and effective healthcare solutions.

2 Related Work

In this section, we discuss the technologies and literature surrounding the development of a patient-friendly overdose detector. First, we describe the implementations of sleep apnea data to model an opioid-OD event. Then, we detail OUD patients' willingness towards such technologies and ongoing work to make devices more patient-centric. Finally, we discuss the current solutions in overdose detection and reversal.

2.1 Hypoxemia Estimation Using ML

Hypoxemia plays a vital role in identifying an overdose due to Opioid-Induced Respiratory Depression (OIRD). As mentioned earlier, the scarcity of clean data pertaining to actual overdose events makes it hard to model [Campbell *et al.*, 2023]. For this reason, we have modeled our hypoxemia classifier around obstructive sleep apnea (OSA) patients experiencing apneic events. Traditional data collection on overdose events faces ethical and practical hurdles, leading us to model our approach on OSA patients. OSA episodes offer parallel biomarker trends to OIRD, albeit with distinctive motion patterns. During an opioid-OD, the individual is completely motionless, and their hypoxemic state tends to increase if the reversal agent naloxone is not administered immediately. Whereas, a patient with OSA can recover from their apneic event during which a reactive body movement is exhibited. This differentiator helps us to use OSA patients as a reasonable model while not misidentifying a person with OSA as having an overdose.

To ensure accurate detection of opioid-OD-induced hypoxemia, we implemented an ML-based approach that utilizes biomarkers estimated from PPG and motion signals. Lazazzera et al. [Lazazzera *et al.*, 2021] developed an apnea and hypopnea classifier that trained on PPG and SpO_2

datasets focusing on changes in PPG amplitude fluctuations(DAP) and oxygen desaturation. Their Fine Gaussian Support Vector Machines model provided a 75.1% accuracy in detecting apneas and hypopneas. Mahmud et al. [Mahmud *et al.*, 2022] focused purely on PPG signals to determine 3 classes of hypoxemia severity: normal, moderate, and critical. The Res-SE-ConvNet deep neural network(DNN) proposed had classified hypoxemia with 96.5% accuracy. Their implementation of convolution neural networks(CNNs) to achieve state-of-the-art accuracy helped incorporate OxyCaps into our classifier to better handle temporal hierarchies. Finally, Hoffman et al. [Hoffman *et al.*, 2022] took a novel approach instead of PPG-based inputs through optical sensors and used a smartphone camera in a hypoxemia study. Their study implemented CNNs to classify hypoxemia levels based on predicted SpO_2 from images collected by the device. They were able to screen hypoxemia as SpO_2 below 90% with an average sensitivity and specificity of 81% and 79%, respectively. It is important to note that these models have sourced their data from clinical establishments and PPG signals captured from the fingertip, a canonical site with high blood perfusion. Our model is built on PPG data sourced from the shoulder, a non-canonical site. Although this results in noisier signals requiring additional filtering, feature engineering, and model tuning, it incorporates OUD patient design needs, thereby increasing willingness to wear the device.

2.2 Patient-Centric Design Principles

Understanding the needs and preferences of the OUD population is crucial when designing life-saving technologies. This section delves into the willingness of OUD patients to adopt overdose monitoring devices and the design considerations that can enhance their acceptance. Ahamad et al. [Ahamad *et al.*, 2019] conducted interviews with 1061 OUD patients, of which 54% were open to wearing an overdose detection device. However, this study did not include tangible prototypes, offering only a general chest-based detection site for consideration. In contrast, Kanter et al. [Kanter *et al.*, 2021b] achieved a higher willingness rate of 76% from their survey of 97 patients. By presenting tangible device prototypes, they provided insights into potential device appearances, placement options, and functionalities. Their findings emphasized that a discreet shoulder-mounted device, capable of monitoring and potentially administering naloxone, would be more favorably received. The shoulder's suitability as a canonical injection site further underscores its potential for integrated injector systems. But Campbell et al. [Campbell *et al.*, 2023] highlighted the challenges in reliably capturing overdose biomarkers using wearables, even under controlled conditions, highlighting the need for additional study.

The importance of patient-centered design cannot be overstated, especially as OUD disproportionately affects marginalized communities. The design must transcend the medical functionality to consider the socio-economic and psychological contexts of the end-users. The Masimo Halo [Knopf, 2023], an FDA-approved device for monitoring the risk of opioid-OD, represents a significant advancement in this field. However, its design, encompassing a smartphone, a home medical hub, and a single-use sensor worn on the wrist and finger, may not fully meet the needs of the economically disadvantaged or those without stable housing. The device's visibility may inadvertently contribute to the stigma associated with OUD, potentially deterring its use among those most in need. Recognizing these limitations, our research aims to focus on improving reusability, affordability, and concealability, ensuring that the device is both accessible and acceptable to all individuals at risk of opioid-OD.

2.3 Opioid Overdose Monitoring

The opioid crisis has spurred significant innovation in sensing methodologies and exploration into additional on-body sites for sensor placement for overdose detection. Nandakumar et al. [Nandakumar *et al.*, 2019] leverages the sonar capabilities of mobile devices to monitor the respiration rate of OUD patients, offering a software-centric solution that capitalizes on patients' existing devices. Dhowan et al. [Dhowan *et al.*, 2019] employs ECG sensors positioned on the shoulder to track respiration rate. This system is integrated with a shoulder device that can administer naloxone from a surgically implanted reservoir beneath the skin. Imtiaz, Bandoian, and Santoro et al. [Imtiaz *et al.*, 2021] shift the focus to SpO_2 detection on the arm, targeting hypoxemia as an indicator rather than respiratory depression. Each of these studies introduces a distinct approach, focusing on a specific biomarker, body location, and hardware configuration. A paramount concern among OUD patients is the inadvertent administration of naloxone or unintentional alerts to bystanders during non-overdose scenarios. In our work, we address this apprehension by leveraging multiple biomarkers for hypoxemia classification. By synthesizing insights from these biomarkers, we aim to reduce false positives, thereby increasing willingness and trust in the technology.

3 Data Collection and Preprocessing

This section outlines the study design, the hardware needed for data collection, and subsequent data refinement. The collected and preprocessed data can be made publicly available [Kumar, 2024]. We first discuss the experimental setup involving sleep apnea patients and the connection between biomarkers of sleep apnea and overdose. Then, we examine the custom hardware developed specifically to acquire biomarker data such as PPG and motion signals from these patients off the shoulder. Finally, we discuss the steps involved in preprocessing, denoising, and feature extraction.

3.1 Study Design

For this experiment, we monitored 19 patients overnight during their scheduled sleep studies at a designated sleep clinic. The study targets sleep apnea patients because of the striking similarities between apneic episodes and opioid-ODs. In both scenarios, patients exhibit motionlessness, ceased breathing, and decreased oxygen saturation [Boyer, 2012a; Berry *et al.*, 2012]. However, OSA patients recover from apneic events within a short period of time while exhibiting resumption of breathing. Motion sensors can capture the resumption of breathing and help differentiate apneic episodes between sleep apnea and opioid-OD.

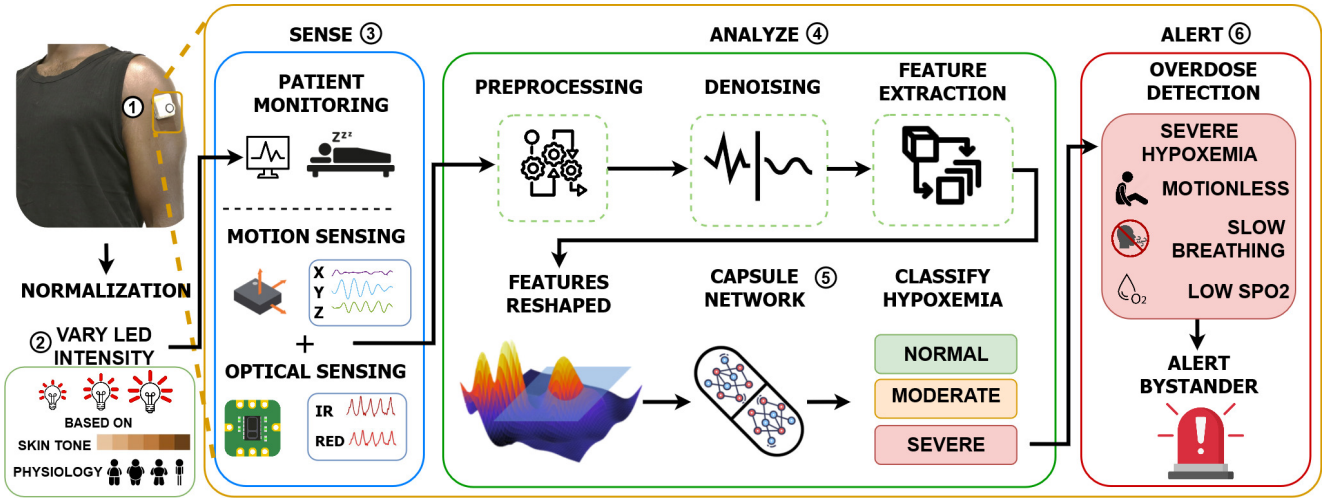

Figure 1: System Architecture (1) Shoulder-based motion & PPG sensor device (2) Normalization of LEDs based on skin tone and physiology (3) Sleep apnea patient data collection (4) Data analysis pipeline (5) OxyCaps hypoxemia classifier (6) Alert mechanism

Prior to the study, we collected demographic and physiological data, including age, race, gender, BMI, and skin tone. Information pertaining to skin tone and physiological details is crucial when evaluating the efficiency of our device's optical normalization algorithm upon initial wear. Previously, Lingamoorthy et al. [Lingamoorthy *et al.*, 2023] outlined the process for optical sensor normalization across various subjects. Each patient was equipped with two pulse oximeter devices for monitoring: our custom-designed shoulder-based device and a commercial FDA-approved finger-based device. Our primary objective was to gather raw biomarker signals through accelerometers and optical sensors, thereby constructing ML models that can estimate hypoxemic states. The commercial device approved by the FDA to collect SpO_2 levels was used to create a labeled dataset. A key focus was on SpO_2 levels, especially drops denoting hypoxemic events, which would simulate conditions resembling an opioid-OD.

3.2 Device Hardware

Detecting opioid-ODs on the shoulder enhances OUD patients' willingness to use a concealable monitoring device [Kanter *et al.*, 2021a]. As such, we designed custom hardware as seen in Figure1. The device was used to collect data and accurately detect biomarkers pertaining to an overdose on a non-canonical site. Our device for this study non-invasively collects PPG and motion data off the shoulders of sleep apnea patients along with a commercial finger pulse oximeter. The oximeter records pulse rate and SpO_2 every second. The oximeter readings are used to label the dataset into 3 stages of hypoxemia [Krejčí *et al.*, 2018], which closely translates to opioid-OD levels.

1. Normal : SpO_2 range 96% to 100%.
2. Moderate : SpO_2 range 92% to 95%.
3. Severe: SpO_2 range 88% to 91%.

The devices use an accelerometer and optical sensor. The 6-axis Inertial Measurement Unit is used to measure 3D accelerometry and gyroscopic data to determine the individual's respiration and motion states. The 2-channel optical sensor collects raw PPG data through reflectance pulse oximetry that is used to estimate oxygen saturation, respiration rate, and motion. The optical sensor uses red and infrared (IR) light-emitting diodes (LED) of wavelengths 660nm and 880nm to illuminate the deltoid and measure the light reflected using a photodetector. The raw red and IR PPG samples, combined with 6-axis motion data, are captured at 25Hz and saved locally in binary format. The device saves LED drive current as an 8-bit intensity value, which adjusts automatically based on the patient's skin tone and physiology. To automatically adjust, a one-time calibration occurs by iterating through light intensities until the maximum perfusion index is reached [Lingamoorthy *et al.*, 2023].

3.3 Preprocessing and Denoising

Data from the test device, collected across all patients, is first decoded and merged into a single file. We then removed samples taken when the patient did not wear either device. This is determined using correlation values between red and IR PPG signals. The commercial device's data is upsampled using a forward fill method due to the infrequent change in SpO_2. Data points with motion artifacts are discarded and are determined by high acceleration magnitude(R_{motion}) from Equation 1.

$$R_{\text{motion}} = \sqrt{x_{\text{axis}}^2 + y_{\text{axis}}^2 + z_{\text{axis}}^2} \tag{1}$$

We filter data points with SpO_2 values below 88% from the commercial device, focusing on the early onset of hypoxemia, typically marked by values below 92% [Lee *et al.*, 2000]. Finally, signals with high noise are discarded based on signal quality index (SQI) calculated using the skewness factor within an 8-second window [Elgendi, 2016]. Temporal discrepancies between the commercial and test devices due to different start times are rectified by aligning the pulse rates from both devices. For PPG signal processing, we apply a 3rd-order Butterworth band-pass filter with cutoff frequencies between 0.5 Hz and 3 Hz to both red and infrared

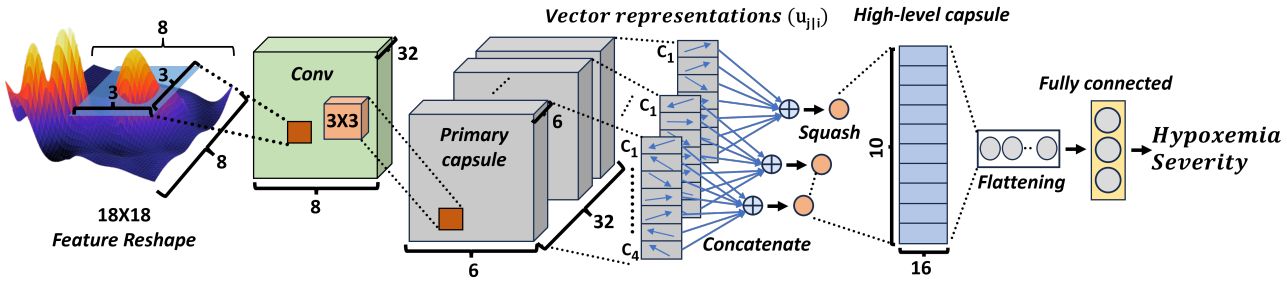

Figure 2: OxyCaps model architecture

wavelengths [Liang *et al.*, 2018] to remove baseline drifts and high-frequency noise. We then segment the data into eight-second epochs, each containing 200 samples [Shuzan *et al.*, 2023]. SpO_2 values are derived as a feature in the form of Ratio of Modulation or (R_{SpO2}) as seen in Equation 2

$$R_{SpO_2} = \left(\frac{AC_{RED}}{DC_{RED}}\right) \Big/ \left(\frac{AC_{IR}}{DC_{IR}}\right) \tag{2}$$

Here AC_{RED}, DC_{RED}, AC_{IR} and DC_{IR} are the AC and DC components of the PPG signal for the red and the infrared wavelengths, respectively [Aguirregomezcorta *et al.*, 2021]. These components are determined by evaluating the standard deviation and mean of each epoch and are used as features for the model. Healthy individuals typically have an SpO_2 range of 95% to 100%. This causes a bias in our dataset, as apneic events, which deviate from this range, only occur sporadically during a patient's sleep study. We downsample based on the least frequent SpO_2 value to address this bias.

4 Methods

In this section, we delve into feature engineering, describe the scaling methods applied to the dataset, and discuss the proposed OxyCaps architecture employed for hypoxemia classification, highlighting their specific hyperparameter settings.

4.1 Features Engineering

Feature engineering plays a pivotal role in crafting a resilient and efficient model. In our study, we generate new features from the raw accelerometer and optical sensor signals. Upon preprocessing the data sourced from sensors, we derive overlaying features pertinent to motion, such as motion levels and respiration metrics. We derive oxygen saturation, respiration features, skin tone, and physiology from PPG signals. Finally, we used Standard scaling to normalize the data due to its consistent handling of outliers compared to other techniques like MinMax or Robust. Our feature processing methods can be seen in our open-source repository [Kumar, 2024].

- *Oxygen Saturation*: pulsatile (AC), non-pulsatile (DC), and the ratio between red and IR signals from PPG
- *Skin Tone and Physiology*: Red and IR LEDs current draw during the normalization process.
- *Motion Levels*: (R_{motion}) from x, y, and z axis accelerometer signals calculate motion intensity.
- *Respiration Features*: Respiration rate, intensity, and duration of inhalation and exhalation.

	Recall			
Features	Normal	Moderate	Severe	Accuracy
w RR	0.90	0.87	0.98	0.92
w/o RR	0.66	0.63	0.81	0.70

Table 1: OxyCaps hypoxemia classification Respiration Rate (RR) feature importance

Respiration features are calculated using optical and motion signals to increase reliability if one signal shows low SQI. PPG signals provide blood volume change [Pimentel *et al.*, 2016] while accelerometry gives us thoraxial motion [Chan *et al.*, 2021]. Combining both sensors, we generate respiration rate, intensity, and duration of inhalation and exhalation through filtering and peak detection algorithms. We also included respiration features time shifted by 10 seconds in the past to track relative changes. These features are vital in increasing hypoxemia classification across all evaluation metrics. Table.1 shows how the OxyCaps model improved recall, the ability to correctly identify classes by 36% for Normal, 38% for Moderate, and 21% for severe hypoxemic classes. The added features also improved the overall accuracy by 31%.

4.2 Proposed Architecture

Our proposed capsule network, OxyCaps, shown in Figure 2, is a variant of the capsule networks proposed in [Sabour *et al.*, 2017]. It has four layers: convolutional layer, primary capsule layer, high-level fully connected capsule layer, and fully connected layer as an output layer. In the first convolutional layer, we utilize a non-linear ReLU activation function, 32 channels, 3×3 kernel size with a stride of 2, and padding set to 0. This layer helps extract salient features from preprocessed features (which are reshaped to (1,18,18)) from optical and motion sensors. Then, the extracted features are fed to the primary capsules layer, whose main role is to replace the scalar-output feature detectors of CNNs with vector-output capsules to preserve the instantiated parameters, such as the local order of features of IR and red PPG signals. It consists of 4 capsules, and each capsule is a convolutional layer with 32 channels, 3×3 kernel size, a stride

of 1, and padding set to 0 and gives an output of $(32 \times 6 \times 6)$, all up, the primary capsules layer gives an output in the form of $(Batch\_size, 4, 32, 6, 6)$. To introduce non-linearity in the primary capsules layer, we used the *Squash* function as defined in Equation 3.

$$Squash(\mathbf{x}) = \frac{\|\mathbf{x}\|^2}{1 + \|\mathbf{x}\|^2} \cdot \frac{\mathbf{x}}{\|\mathbf{x}\|} \tag{3}$$

The penultimate layer has a single 16-D capsule and receives an input of $(Batch\_size, 4, 32, 6, 6)$ from the primary capsules layer and provides an output $(Batch\_size, N_c, 16)$. Primarily, each capsule in the penultimate layer, whose length of the activity vector designates the compressed form of motion and sensor-based features, receives input from all the capsules in the layer below. Here, the N_c represents the number of capsules in the penultimate layer, which is set to 10 (found during the tuning of the model). Then, we apply flattening to the receiving input $(Batch\_size, N_c, 16)$ and feed to the output layer having three neurons with a linear transformation to predict 3 stages of hypoxemia.

As mentioned in [Sabour *et al.*, 2017], the capsule network permits the networks to automatically learn child-parent (or part-whole) relationships. In the hypoxemia classification task, different samples from a patient with the same SpO_2 category are supposed to have similar skin tones but with different features. In this paper, we have done transformation matrices to create a prediction vector (vote) $\hat{u}_{j|i}$ from its child capsule i to the parent capsule j. The first one shares a weight matrix (W_{ij}) between each child capsule (u_i) and parent capsule (v_j). Properly, each corresponding vote can be calculated by:

$$\hat{u}_{j|i} = W_{ij} u_i + \hat{b}_{j|i} \in R^d \tag{4}$$

where $\hat{b}_{j|i}$ is a capsule bias term.

In the architecture, the routing is done between the primary capsule layer and the high-level fully connected capsule layer. The main idea of using dynamic routing is to build a non-linear map in a repetitious way to confirm that the output of each capsule is directed to its corresponding parent in the subsequent layer. Also, the connection strength can be increased or decreased between capsules via dynamic routing, which is more efficacious with respect to earliest routing strategies like max-pooling in CNN, which detects whether a feature is present in any position but loses spatial information about the feature. In our case, we have used the same routing algorithm as proposed [Sabour *et al.*, 2017]. The capsules in the penultimate layer below are flattened and fed into a fully connected output layer containing three neurons with a linear transformation to predict hypoxemia severity. In our experiment, all the routing logits (b_{ij}) are initialized to zero, and the number of iterations is kept to 3. Our implementation is in Pytorch [Paszke *et al.*, 2019] with a batch_size of 512, epochs set to 101, AdamW optimizer (learning rate 1e-3) with a cross-entropy loss function.

5 Results and Evaluation

In this section, we assess existing ML models that have traditionally been employed for hypoxemia detection using only PPG data from canonical sites [Mahmud *et al.*, 2022; Hoffman *et al.*, 2022; Lazazzera *et al.*, 2021]. We extend this evaluation to newer sites, like the shoulder, and compare the performance of other ML models against OxyCaps.

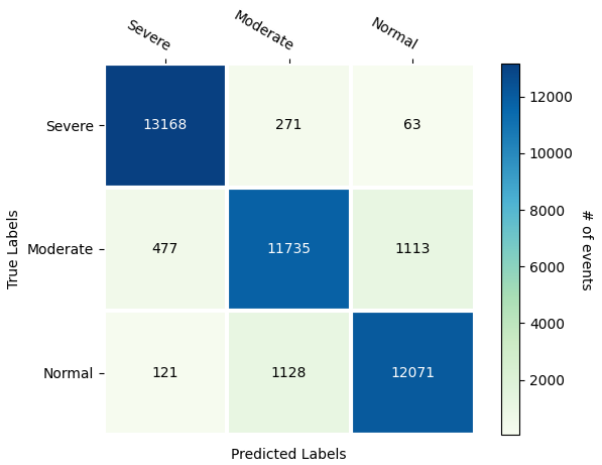

Figure 3: OxyCaps confusion matrix

5.1 Evaluation Metrics

To validate the performance of our hypoxemia classifier, we use a suite of evaluation metrics, each offering a unique perspective on the model's performance. This section delves into the specifics of each metric and its significance in the context of our study. The metrics used are Precision, Recall, F1-Score, and Accuracy.

- *Precision*: Assesses reliability in predicting hypoxemia stages, reducing false alarms.
- *Recall*: Measures capability to identify all hypoxemia instances, minimizing missed detections.
- *F1-Score*: Compares classifiers, balancing false positives and negatives, important for imbalanced classes.
- *Accuracy*: Evaluates overall correct classification.

5.2 Cross-validation

To assess the resilience of our hypoxemia classifier, we compare our OxyCaps methodology with existing models and employ a 5-fold cross-validation, mirroring the approach taken in other approaches [Lazazzera *et al.*, 2021; Mahmud *et al.*, 2022]. This method provides a comprehensive evaluation by partitioning the dataset into subsets and iteratively training and testing the model. We adopted this strategy to optimize the use of our training data and to prevent overfitting.

Next, we discuss the performance of the OxyCaps model as compared to other ML models and approaches. Empirical studies on models—ranging from RandomForest to advanced neural networks as seen in Table.2. These models, however, did not match the performance of OxyCaps, which shows almost 7% better handling of hypoxemia classification. OxyCaps shows proficiency in preserving the spatial hierarchy of features, which helps interpret complex physiological signals. Its vector-based capsules effectively encode the probability

Model	Precision	Recall	F1	Accuracy
RandomForest	0.88	0.88	0.88	0.88
ExtraTrees	0.88	0.88	0.88	0.88
LightGBM	0.84	0.84	0.84	0.84
XGBoost	0.88	0.88	0.88	0.88
CatBoost	0.85	0.85	0.85	0.86
HistGradient	0.89	0.89	0.89	0.90
3-Layers ANN	0.89	0.89	0.89	0.89
OxyCaps	**0.92**	**0.92**	**0.92**	**0.92**

Table 2: Empirical study 5-Fold cross-validation

Class	Precision	Recall	F1-Score
Normal	0.91	0.91	0.91
Moderate	0.89	0.88	0.89
Severe	0.96	0.98	0.97

Table 3: OxyCaps hypoxemia classifier 5-Fold CV

and properties of features, such as motion patterns from accelerometry and individual physiological signatures, which are pivotal for hypoxemia classification. OxyCaps can accurately classify levels of hypoxemia with 92% accuracy off the shoulder while maintaining high sensitivity towards false positives is critical for determining an opioid-OD. Ensuring our false positive rates are low is important to maximize willingness to wear an overdose alert device.

The other ML [Mahmud *et al.*, 2022; Hoffman *et al.*, 2022; Lazazzera *et al.*, 2021] based approaches that were discussed above incorporate PPG signals to classify hypoxemia. Lazazzera et al. [Lazazzera *et al.*, 2021] developed an apnea and hypopnea classifier using Fine Gaussian Support Vector Machines with an accuracy of 75.1%. The Res-SE-ConvNet deep neural network (DNN) developed by Mahmud et al. [Mahmud *et al.*, 2022] had a 3-class hypoxemia classifier accuracy of 96.5%. Finally, the smartphone implementation by Hoffman et al. [Hoffman *et al.*, 2022] was able to screen hypoxemia as SpO_2 below 90% with an average sensitivity and specificity of 81% and 79%, respectively. OxyCaps performs reasonably well compared to these models, as seen in Table.3. It is important to note that these methods used a different range for classifying hypoxemia. Furthermore, these approaches were developed for finger-based sensors but do not translate directly for a shoulder-mounted sensor. This is mainly due to variations in perfusion, skin tones, physiology, and signal-to-noise ratio. This variation of the detection site introduces complexity, leading to an uneven comparison when evaluating other datasets in our model.

5.3 Limitations and Future Work

Our study leveraged the OxyCaps model for hypoxemia classification from PPG and motion signals detected off the shoulder. While OxyCaps's dynamic routing offers unique advantages, it introduces computational complexity, making it more time-consuming to train than Convolutional Neural Networks (CNNs). The data collection posed its own set of challenges. We utilized a custom-developed pulse oximeter, which, although specific to our needs, is not FDA-grade and is likely to be more noise-sensitive. Additionally, while FDA cleared, the commercial pulse oximeter is not ICU grade, potentially introducing variability in the SpO_2 ground truth. The dataset, derived from only 19 patients, does not fully represent the broader population's PPG signal variability across different skin tones and body types. The shoulder as a detection site influences PPG signal quality and subsequent hypoxemia detection due to its inherently high signal-to-noise ratio and the deltoid's low perfusion, which pose challenges.

Several advancements can enhance the accuracy of our classifier. Firstly, we calibrate our shoulder-based detection device in a controlled hypoxemia lab that provides Drug-Induced Apnea Testing. By doing so, we can train our model against the gold standard arterial blood oxygen (SaO_2) values, ensuring its precision in critical scenarios. Secondly, we aim to provide take-home devices for OUD patients, allowing us to collect PPG and motion data directly from the target demographic. This real-world data can provide insights into the unique challenges and nuances of overdose estimation in the OUD community. Lastly, efforts will be directed towards enhancing the sensor's hardware capabilities. By improving the signal-to-noise ratio and incorporating multi-channel spectral sensing, we aim to capture cleaner physiological data, creating a comprehensive and reliable detection system.

Due to the number of low apneic events, a 5-fold validation approach had to be performed to analyze the model performance instead of patient disjoint evaluation. Moreover, excluding a certain number of patients' data and isolating them only for test purposes severely affects the model training as it will not have enough unique training samples for the severe hypoxemia class. Therefore, data collected in a controlled environment with a balanced distribution will be employed to verify the model using patient hold-out tests.

6 Conclusion

Accurate hypoxemia detection from non-traditional sites is pivotal for the evolution of patient-centric opioid-OD alert systems. Such advancements not only enhance the precision of alert mechanisms but also pave the way for integrating opioid-OD reversal devices, especially given the shoulder's status as a standard injection site for drugs like naloxone. Historically, SpO_2 for in-hospital hypoxemia detection has been predominantly centered around regions like the fingertip, toe, forehead, and earlobe, primarily due to their high perfusion index and favorable signal-to-noise ratio. However, this work introduces a paradigm shift by proposing a capsule network-based methodology for hypoxemia detection, leveraging spatial relationships between features while considering temporal delays. This innovative approach capitalizes on distinctive features encompassing skin tone, individual physiology, and motion dynamics, ensuring a comprehensive evaluation in real-world contexts. Significantly, our methodology demonstrates high accuracy in estimating severe hypoxemic states observed during an overdose from the shoulder. This endeavor not only pushes the boundaries of opioid-OD detection but also underscores the potential of leveraging unconventional sites for critical medical applications.

Ethics Statement

The study followed strict ethical standards and was approved by the Institutional Review Board (IRB protocol number 849910). All procedures involving human participants were in accordance with the ethical standards of the institutional and national research committees. We ensured the confidentiality and anonymity of all participants involved in the study. Additionally, ethical considerations concerning the use of technology in sensitive health scenarios were addressed, emphasizing user consent and data protection.

Acknowledgements

We gratefully acknowledge the support and contributions of Aaron Wang, Xiaonan Ma, Ayan Mandal, Korey Henderson, and Ethan Donlon for their assistance in data collection. Special thanks to the participants and staff at the Penn Sleep Center for their cooperation and support throughout this study. We also thank the Drexel University Coulter Translational Research Partnership Program and the NIH/NIDA for funding this research under STTR grant R41DA056276. Additionally, we disclose that Anush Lingamoorthy, a co-author of this paper, holds equity in AltruMed, a medical device company that is in the process of commercializing technology related to the research presented here. This relationship has been managed in accordance with the ethical guidelines of Drexel University.

References

[Aguirregomezcorta *et al.*, 2021] Idoia Badiola Aguirregomezcorta, Vladimir Blazek, Steffen Leonhardt, and Christoph Hoog Antink. Learning about reflective ppg for spo2 determination using machine learning. *Current Directions in Biomedical Engineering*, 7(2):33–36, 2021.

[Ahamad *et al.*, 2019] Keith Ahamad, Huiru Dong, Cheyenne Johnson, Kanna Hyashi, Kora DeBeck, MJ Milloy, and Evan Wood. Factors associated with willingness to wear an electronic overdose detection device. *Addiction science & clinical practice*, 14(1):1–5, 2019.

[Berry *et al.*, 2012] Richard B Berry, Rita Brooks, Charlene E Gamaldo, Susan M Harding, C Marcus, Bradley V Vaughn, et al. The aasm manual for the scoring of sleep and associated events. *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, 176:2012, 2012.

[Boyer, 2012a] Edward W Boyer. Management of opioid analgesic overdose. *New England Journal of Medicine*, 367(2):146–155, 2012.

[Boyer, 2012b] Edward W Boyer. Management of opioid analgesic overdose. *New England Journal of Medicine*, 367(2):146–155, 2012.

[Campbell *et al.*, 2023] Anne Campbell, Sharon Millen, Li Guo, Uisce Jordan, Amanda Taylor-Beswick, Chris Rintoul, and Aisling Diamond. Reducing opioid related deaths for individuals who are at high risk of death from overdose: a co-production study with people housed within prison and hostel accommodation during covid-19. *Frontiers in Public Health*, 11:1080629, 2023.

[Chan *et al.*, 2021] Justin Chan, Vikram Iyer, Anran Wang, Alexander Lyness, Preetma Kooner, Jacob Sunshine, and Shyamnath Gollakota. Closed-loop wearable naloxone injector system. *Scientific reports*, 11(1):1–13, 2021.

[Chang *et al.*, 2018] Hsien-Yen Chang, Hadi Kharrazi, Dave Bodycombe, Jonathan P Weiner, and G Caleb Alexander. Healthcare costs and utilization associated with high-risk prescription opioid use: a retrospective cohort study. *BMC medicine*, 16(1):1–11, 2018.

[Dhowan *et al.*, 2019] Bahar Dhowan, Jongcheon Lim, Michael D MacLean, Alycia G Berman, Min Ku Kim, Qi Yang, Jacqueline Linnes, Chi Hwan Lee, Craig J Goergen, and Hyowon Lee. Simple minimally-invasive automatic antidote delivery device (a2d2) towards closed-loop reversal of opioid overdose. *Journal of Controlled Release*, 306:130–137, 2019.

[Elgendi, 2016] Mohamed Elgendi. Optimal signal quality index for photoplethysmogram signals. *Bioengineering*, 3(4):21, 2016.

[Hoffman *et al.*, 2022] Jason S Hoffman, Varun K Viswanath, Caiwei Tian, Xinyi Ding, Matthew J Thompson, Eric C Larson, Shwetak N Patel, and Edward J Wang. Smartphone camera oximetry in an induced hypoxemia study. *NPJ Digital Medicine*, 5(1):146, 2022.

[Imtiaz *et al.*, 2021] Mohammad S Imtiaz, Charles V Bandoian, and Thomas J Santoro. Hypoxia driven opioid targeted automated device for overdose rescue. *Scientific Reports*, 11(1), 2021.

[Kanter *et al.*, 2021a] Katie Kanter, Ryan Gallagher, Feyisope Eweje, Alexander Lee, David Gordon, Stephen Landy, Julia Gasior, Haideliza Soto-Calderon, Peter F Cronholm, Ben Cocchiaro, et al. Willingness to use a wearable device capable of detecting and reversing overdose among people who use opioids in philadelphia. *Harm reduction journal*, 18(1):1–14, 2021.

[Kanter *et al.*, 2021b] Katie Kanter, Ryan Gallagher, Feyisope Eweje, Alexander Lee, David Gordon, Stephen Landy, Julia Gasior, Haideliza Soto-Calderon, Peter F Cronholm, Ben Cocchiaro, et al. Willingness to use a wearable device capable of detecting and reversing overdose among people who use opioids in philadelphia. *Harm Reduction Journal*, 18(1):1–14, 2021.

[Knopf, 2023] Alison Knopf. Masimo "halo" gets fda go-ahead for otc and rx. *Alcoholism & Drug Abuse Weekly*, 35(16):4–5, 2023.

[Krejčí *et al.*, 2018] Jakub Krejčí, Michal Botek, and Andrew J McKune. Dynamics of the heart rate variability and oxygen saturation response to acute normobaric hypoxia within the first 10 min of exposure. *Clinical physiology and functional imaging*, 38(1):56–62, 2018.

[Kumar, 2024] Abhishek Kumar. Drug overdose vital-signs evaluator using machine learning. https://github.com/abhishekkumarm98/

Drug-Overdose-Vital-Signs-Evaluator, 2024. Accessed: 2024-06-26.

[Lazazzera *et al.*, 2021] Remo Lazazzera, Margot Deviaene, Carolina Varon, Bertien Buyse, Dries Testelmans, Pablo Laguna, Eduardo Gil, and Guy Carrault. Detection and classification of sleep apnea and hypopnea using ppg and spo_2 signals. *IEEE Transactions on Biomedical Engineering*, 68(5):1496–1506, 2021.

[Lee *et al.*, 2000] Wally W. Lee, Kurt Mayberry, Robert Crapo, and Robert L. Jensen. The accuracy of pulse oximetry in the emergency department. *The American Journal of Emergency Medicine*, 18(4):427–431, 2000.

[Liang *et al.*, 2018] Yongbo Liang, Mohamed Elgendi, Zhencheng Chen, and Rabab Ward. An optimal filter for short photoplethysmogram signals. *Scientific data*, 5(1):1–12, 2018.

[Lingamoorthy *et al.*, 2023] Anush Lingamoorthy, Amanda Watson, Korey Henderson, Ayan Mandal, David Gordon, Xiaonan Ma, James Weimer, Nagarajan Kandasamy, and Jacob S Brenner. Dove: Shoulder-based opioid overdose detection and reversal device. In *2023 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pages 56–67. IEEE, 2023.

[Mahmud *et al.*, 2022] Talha Ibn Mahmud, Sheikh Asif Imran, and Celia Shahnaz. Res-se-convnet: A deep neural network for hypoxemia severity prediction for hospital inpatients using photoplethysmograph signal. *IEEE Journal of Translational Engineering in Health and Medicine*, 10:1–9, 2022.

[Nandakumar *et al.*, 2019] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Jacob E Sunshine. Opioid overdose detection using smartphones. *Science translational medicine*, 11(474):eaau8914, 2019.

[Ogeil *et al.*, 2018] Rowan P Ogeil, Jeremy Dwyer, Lyndal Bugeja, Cherie Heilbronn, Dan I Lubman, and Belinda Lloyd. Pharmaceutical opioid overdose deaths and the presence of witnesses. *International Journal of Drug Policy*, 55:8–13, 2018.

[Paszke *et al.*, 2019] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[Pimentel *et al.*, 2016] Marco AF Pimentel, Alistair EW Johnson, Peter H Charlton, Drew Birrenkott, Peter J Watkinson, Lionel Tarassenko, and David A Clifton. Toward a robust estimation of respiratory rate from pulse oximeters. *IEEE Transactions on Biomedical Engineering*, 64(8):1914–1923, 2016.

[Sabour *et al.*, 2017] Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic routing between capsules. *Advances in neural information processing systems*, 30, 2017.

[Shuzan *et al.*, 2023] Md Nazmul Islam Shuzan, Moajjem Hossain Chowdhury, Muhammad EH Chowdhury, Murugappan Murugappan, Enamul Hoque Bhuiyan, Mohamed Arslane Ayari, and Amith Khandakar. Machine learning-based respiration rate and blood oxygen saturation estimation using photoplethysmogram signals. *Bioengineering*, 10(2):167, 2023.