International Journal of Data Science and Analytics (2025) 20:4261–4272
https://doi.org/10.1007/s41060-025-00716-z

REGULAR PAPER

Robust online drop detection for cochlear implant surgery

Yuansan Liu[1] · Sudanthi Wijewickrema[2] · Christofer Bester[2] · Stephen O'Leary[2] · James Bailey[1]

Received: 19 September 2024 / Accepted: 2 January 2025 / Published online: 10 January 2025
© The Author(s) 2025

Abstract
Hearing loss is an impairment that affects a large population worldwide. Cochlear implant (CI) surgery is an effective solution, especially in patients with residual, natural hearing. However, the risk of intracochlear damage during CI surgery, which impairs residual hearing, remains as a concern. It has been shown that intra-operative monitoring (using electrocochleography, specifically the cochlear microphonic) during CI surgery can detect drops that are associated with such trauma, resulting in its potential prevention. Researchers have attempted to automate the process of drop detection by implementing supervised models. However, these have so far failed to achieve good reliability and generalizability. To address these issues, we develop an unsupervised model based on the Bayesian method with a scoring function derived from the logistic function. The proposed model, as a data-driven method, directly learns from data without the need for any supervision which enhances its adaptability to diverse patient populations. Empirical results demonstrate the proposed model's high accuracy with a low false alarm rate (18 extra detection against 146 in previous work), as well as its robustness when processing new data. This indicates its potential for real-world application to improve safety during CI surgery.

Keywords Bayesian method · Online change point detection · Cochlear implantation · Data-driven method

1 Introduction

Hearing loss, as a critical impairment, has been ranked as the third largest cause of disabilities by the latest Global Burden of Disease study [1]. It can affect not only patients' communication abilities but also their overall life quality. Cochlear implant (CI) surgery has emerged as an effective solution to hearing loss, especially for patients that have retained some residual, natural hearing [2, 3].

✉ James Bailey
baileyj@unimelb.edu.au

Yuansan Liu
yuansanl@student.unimelb.edu.au

Sudanthi Wijewickrema
sudanthi.wijewickrema@unimelb.edu.au

Christofer Bester
christofer.bester@unimelb.edu.au

Stephen O'Leary
sjoleary@unimelb.edu.au

[1] School of Computing and Information Systems, The University of Melbourne, Melbourne 3053, VIC, Australia

[2] Department of Surgery (Otolaryngology), The University of Melbourne, Melbourne 3053, VIC, Australia

However, one major barrier to patients receiving this treatment is potential intracochlear damage during surgery. This has been shown to lead to irreversible loss of any residual acoustic hearing, which has been demonstrated to play an important role in the development of post-operative speech perception [4, 5]. As such, in the past few decades, increasing attention has been paid to the preservation of residual hearing during CI surgery. One way to achieve this goal is intra-operative monitoring (and subsequent intervention), typically through electrocochleography (ECochG). Previous studies have shown that a 30% drop in amplitude of one ECochG component, namely the "cochlear microphonic" (CM), can reflect intracochlear damage [6, 7]. Therefore, we refer to these as "traumatic drops" and focus on their detection to subsequently prevent intra-operative trauma.

This rule of thumb of a 30% drop has been used as a guideline during CI surgery to detect traumatic drops and prevent further damage through immediate intervention [8, 9]. Furthermore, researchers have also developed supervised machine learning algorithms to automate the process of drop detection [10]. Although their models achieve good performance with respect to sensitivity, several challenges remain to be addressed to build a reliable and efficient method: (1) supervised machine learning methods typically require large

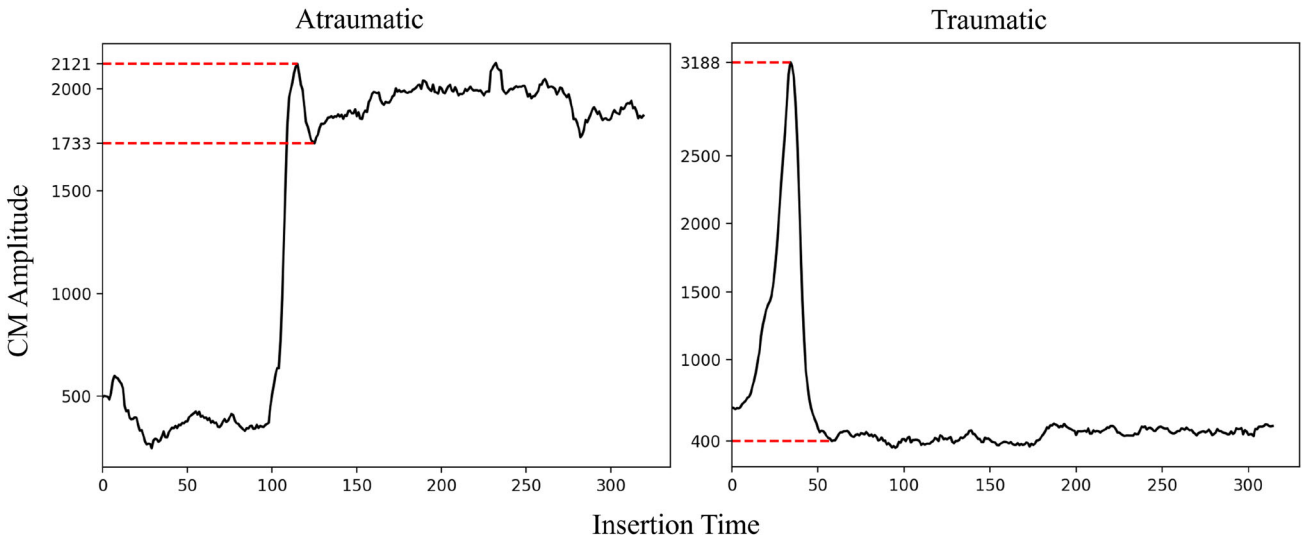


Fig. 1 CM amplitudes for two different types of CI surgeries: atraumatic (left) and traumatic (right)

amounts of data, which is not always practical to collect in real-world applications such as surgery, (2) the variance between patient data is high, mostly due to the different levels of residual hearing, which poses a crucial obstacle for achieving generalizability, and (3) a surgical model needs to handle the precision-recall trade-off well: an over-sensitive model can detect many false positives, which can overwhelm the surgeon, while a model that detects more false negatives will not be accurate enough.

Motivated by the aforementioned challenges, we propose an unsupervised data-driven model based on the Bayesian method and the logistic function to learn each patient's data individually. Specifically, we divide this problem into two sub-problems: change point detection and drop detection. We first utilize a Bayesian online change point detection method to find drop phases in the data of each patient. Then, we develop a scoring function to assign a probability score to each point inside the drop phases based on the 30% drop rule. We call the model **Bayesian logistic online drop detector (BLODD)**. Consequently, as this is an unsupervised model that learns from observations of each patient, training is not required, which alleviates the need for large amounts of historical data. Empirical results demonstrate that the model significantly reduces the number of false alarms while maintaining the sensitivity at a high level, which reflects the efficiency and effectiveness of the model.

2 Background and previous work

2.1 CI surgery and ECochG data

CI surgery is unlike the use of hearing aids that amplify sound. It involves the surgical insertion of an electrode array into the inner ear to compensate for malfunctions and delivers electrical stimulation to the auditory nerve. In addition, most electrodes can also record the inner ear response (ECochG), to enable real-time monitoring of the insertion process. The recordings presented in this work were taken from patients included in a prospective study of ECochG patterns during cochlear implantation at the Royal Victorian Eye and Ear Hospital, Melbourne, under Human Research Ethics Committee review of the Royal Victorian Eye and Ear Hospital (#14/1171 H). Signals were recorded during implantation of Cochlear Limited's Nucleus straight CI422 or CI522 arrays. Patients had pre-operative audiometric thresholds of 80 dB HL or better at 500-Hz. These recordings used the University of Melbourne's ECochG system, Cochlear Response Telemetry (CRT), that records intra-operative ECochG using the apical electrode on the array in response to a 100-dB HL pure tone of 6-ms duration, with a recording window of 12-ms, at 20-kHz. Data were manually labeled according to the experience of an auditory neuroscientist [7]. The ECochG data consist of two components, from hair cells and from auditory neurons, which are later transformed into "cochlear microphonic" (CM) and "auditory nerve neurophonic" (ANN), respectively. Previous work studied the relationship between these responses and post-operative hearing preservation and observed that a 30% drop in CM amplitude leads to lower preservation of residual hearing and this has been used during surgery as a guideline to reduce trauma [6, 11]. Figure 1 shows CM amplitudes during two different CI surgeries: one atraumatic (left) and the other traumatic (right).

This monitoring process was traditionally carried out by human experts, but recent work [10] has implemented and compared several machine learning algorithms on automatic detection of these traumatic drops. This work converted the

original problem into a classification problem: each time-point in a patient was considered an instance with assigned features, and assigned a label that represented its state (atraumatic as 0 and traumatic as 1). Then, classifiers such as decision tree or support vector machine were implemented to predict the labels for each point. However, the traumatic points were less represented among all patient data, which made a supervised model easily biased toward the over-represented class. To mitigate this problem, they manually re-weighted the classes based on the training set class ratio. Additionally, they also proposed a post-processing step which corrected some false predictions based on pre-defined rules. They found the adaptive boosting tree (AdaBoost tree) performed best with their proposed processing steps, which achieved 95.6% accuracy and 85.8% recall. Although their method can increase the model's sensitivity, several issues still remain. Firstly, a supervised model inherently relies on its training process and requires a sufficient amount of data to capture common properties. However, considering the real-world properties of data in the domain of surgery, this may not always be practical. Secondly, the data between patients vary considerably, depending on factors such as the level of residual hearing. As such, it is difficult to learn general rules from training data and apply them to new cases. Thirdly, manual adjustment of class weights requires knowledge of the data distribution prior to training, which may not always be possible. These concerns may lead to the development of an over-sensitive model. For example, the AdaBoost Tree proposed in previous work achieves a high recall score at the cost of precision (85.8% and 43.6%, respectively). This indicates that a large amount of false alarms will be produced by this model upon deployment, degrading the reliability of the model.

2.2 Change point detection

Change point detection (CPD) is the problem of detecting time points where the properties of the time series change abruptly [12]. A change point occurs when one state ends and another different state starts. The objective of CPD is to identify the parameter changes behind these and subsequently detect the change points. These change points help in analyzing the properties of each state and the entire time series. This work focuses on drop states, especially the traumatic ones. Figure 2 shows an example of CPD on CM amplitude.

These data contain 4 change points (red dashed lines). The intervals between these change points are distinct states characterized by consistent statistical properties (e.g., rising, dropping, or stationary). We emphasize the drop states and highlight the significant traumatic drop points of interest.

Online change point detection (OCPD) is the real-time variant of regular CPD, where the detection is carried out while data is being received. This online detection increases

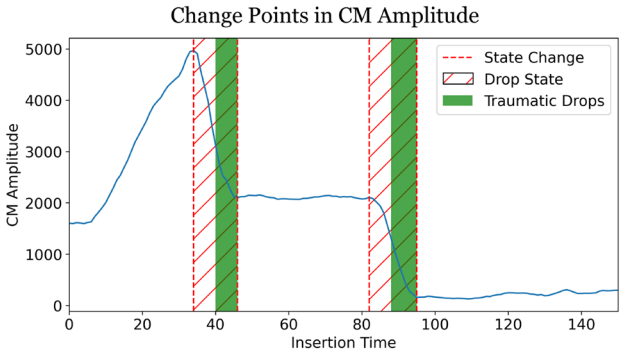


Fig. 2 Example of change point detection in CM amplitude

the complexity of regular CPD, as future information is unavailable during the detection process.

OCPD has been studied for several decades, and many machine learning (ML) methods, mostly statistical models, have been proposed to solve it. One classic solution is to consider the probability density of two data sub-sequences, where a change occurs if they differ significantly. The most efficient approach is to directly estimate the density ratio, which avoids the estimation of the density and makes the problem easier to solve [13, 14]. Relative unconstrained least-squares importance fitting (RuLSIF) uses a relative Pearson divergence to estimate the density ratio directly and achieves great success [15]. However, this kind of method cannot be implemented in real time because two consecutive intervals are needed to estimate and compare the densities. Another type of statistical formulation is probabilistic models. The Bayesian family is a typical representative. These apply Bayesian inference and use the posterior distribution to detect changes. Bayesian online change point detection (BOCPD) introduces the 'run length' variable and enables the real-time detection of change points with the Bayesian approach [16]. Gaussian processes (GP), which are similar to BOCPD, use the Gaussian distribution to do one-step-ahead prediction and determine the changes based on its prediction [17]. The difference from BOCPD is that GP uses the entire history of observations for prediction, while BOCPD uses observations from the last change point, which can improve the effectiveness and the efficiency of the inference.

3 Methodology

To address the challenges inherent in previous work, an unsupervised data-driven approach is proposed. In contrast to [10], we view traumatic drop detection problem as a special OCPD problem. Although standard OCPD methods cannot be directly applied to our task (because data properties might not always change around the 30% threshold, our guideline for traumatic drops), solutions can be developed based on

them. We divide the entire task into two steps: First, we apply the Bayesian online CPD (BOCPD) algorithm to detect drop phases in real time, because it *(1) does not need any training, (2) inferences are based on data, (3) enables each patient's data to be modeled individually, (4) incorporates different types of prior knowledge properly, enabling it to handle situations where prior information is limited, and (5) can detect changes in real time* [16, 18]. We validate this selection by comparing BOCPD with two other OCPD methods (RuLSIF and GP) on a surgical drop detection task. Also, a scoring function is developed to assign probability scores to each point inside drop phases to detect traumatic ones (≥ 0.5).

3.1 Bayesian online change point detection

The first part of our model is to detect drop phases. We directly apply BOCPD [16] which has achieved success in solving the online CPD problem. In this method, the run length r_t is used to model the data since the last change point and takes two value types: either incremental $r_{t-1}+1$ when there is no change to the current state, or 0 when a change occurs. The objective is to infer the posterior for both the run length and data. Given a sequence of observations $x_1, x_2, \ldots, x_t$ (denoted as $\boldsymbol{x}_{1:t}$), the run length posterior can be written as:

$$P(r_t|\boldsymbol{x}_{1:t}) = \frac{P(r_t, \boldsymbol{x}_{1:t})}{P(\boldsymbol{x}_{1:t})} \tag{1}$$

the joint distribution over run length and observations are calculated recursively:

$$P(r_t, \boldsymbol{x}_{1:t}) = \sum_{r_{t-1}} P(r_t|r_{t-1}) P(x_t|r_{t-1}, \boldsymbol{x}_t^{(r)}) P(r_{t-1}, \boldsymbol{x}_{1:t-1}) \tag{2}$$

where $P(x_t|r_{t-1}, \boldsymbol{x}_t^{(r)})$ is the data posterior, $\boldsymbol{x}_t^{(r)}$ denotes the set of observations associated with the corresponding run length r_t. The initial state of this recursion is $P(r_0 = 0) = 1$ which assumes the first observation is always the start of a new state.

The prior distribution of the run length $P(r_t|r_{t-1})$ encodes background knowledge about the change point, and thus is also called the "change point prior". Let T be a random variable denoting the time of change associated with the current run length τ. We can define a hazard function $H(\tau)$ as:

$$H(\tau) = \frac{P_{gap}(T = \tau)}{\sum_{t=\tau}^{\infty} P_{gap}(T = t)} \tag{3}$$

where P_{gap} represents that the probability of the time interval ('gap') between two change points is τ. The hazard function $H(\tau)$, derived from P_{gap}, qualifies the probability of a change point happening at current time step given that it has not occurred in previous τ steps.

Using the hazard function $H(\tau)$, $P(r_t|r_{t-1})$ can be written as:

$$P(r_t|r_{t-1}) = \begin{cases} H(r_{t-1}+1), & \text{if } r_t = 0 \\ 1 - H(r_{t-1}+1), & \text{if } r_t = r_{t-1}+1 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

When T is a discrete geometric random variable with success rate $1/\lambda$, the hazard function becomes a memory-less process and we can write it as $H(\tau) = 1/\lambda$. We directly use this constant hazard function in our approach.

The data posterior $P(x_t|r_{t-1}, \boldsymbol{x}_t^{(r)})$ is inferred by a probabilistic model, which leverages the convenience of the exponential family and their conjugate priors. The choice of model depends on the background knowledge of the data [16]. We use the *Gaussian* distribution with unknown mean and variance to avoid the need for informative prior information which is hard to obtain in surgical data and which thus increases the generalizability of the model. The conjugate prior in this case is *Normal-Gamma* [19]:

$$NG(\mu, \lambda) = \mathcal{N}(\mu|\mu_t, (\kappa_t\lambda)^{-1}) Ga(\lambda|\alpha_t, \beta_t) \tag{5}$$

where $\mathcal{N}(\cdot)$ is a Gaussian distribution and $Ga(\cdot)$ is a gamma distribution. The parameters are $\mu_t, \kappa_t, \alpha_t$, and β_t.

The corresponding posterior is a T distribution:

$$P(x_t|r_{t-1}, \boldsymbol{x}_t^{(r)}) = t_{2\alpha}\left(\tilde{\boldsymbol{x}}|\mu_{t+1}, \frac{\beta_{t+1}(\kappa_{t+1}+1)}{\alpha_{t+1}\kappa_{t+1}}\right) \tag{6}$$

the posterior parameters are updated given a new observation:

$$\mu_{t+1} = \frac{\mu_t\kappa_t + x_{t+1}}{\kappa_t + 1} \tag{7}$$

$$\kappa_{t+1} = \kappa_t + 1 \tag{8}$$

$$\alpha_{t+1} = \alpha_t + \frac{1}{2} \tag{9}$$

$$\beta_{t+1} = \beta_t + \frac{\kappa_t(x_{t+1} - \mu_t)^2}{2(\kappa_t + 1)} \tag{10}$$

Choosing proper initial values for these parameters is the most challenging aspect in this Bayesian method. Following the discussion in [20], weakly informative priors show optimal performance compared to completely non-informative ones. As such, we select values that make the prior knowledge weakly informative. To this end, we first set $\mu_0 = 0, \kappa_0 = 1$ assuming that the base case of the data follows a standard normal distribution. The most important parameters are α_0 and β_0 in the *Gamma* distribution which directly impact the informativeness of the prior belief. We set $\alpha_0, \beta_0 = 0.1$ to allow regularization upon $\mathcal{N}(0, 1)$ while avoiding the prior

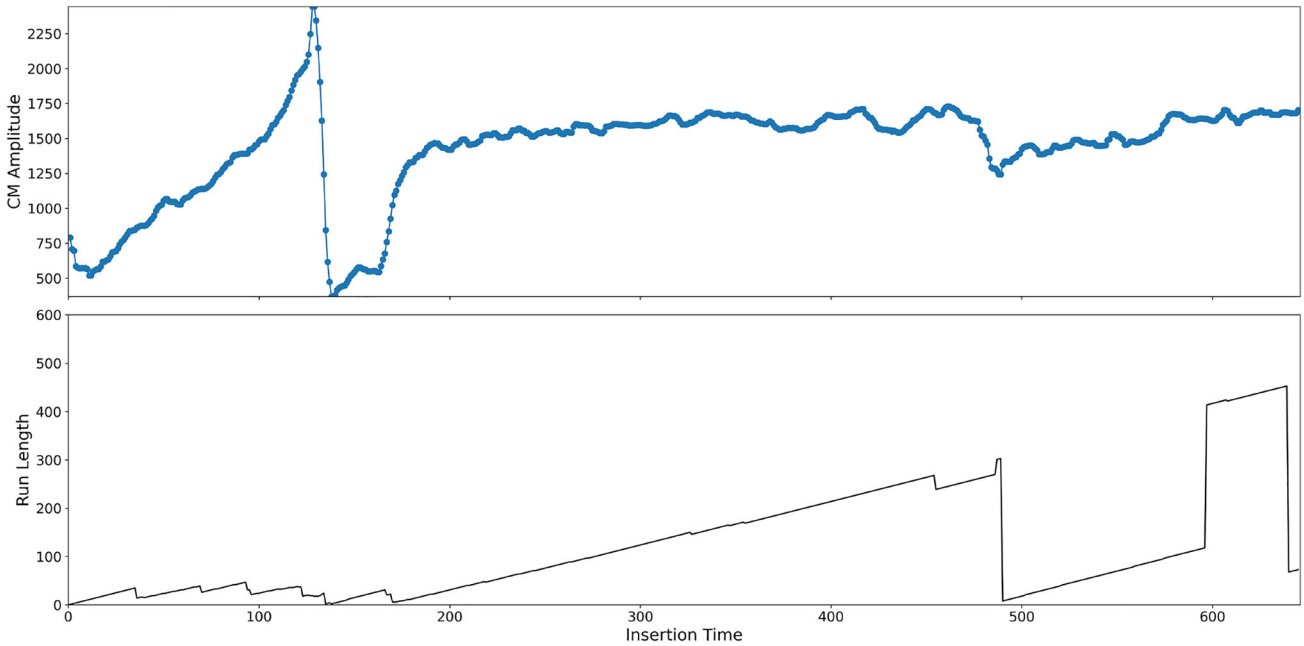


Fig. 3 BOCPD with CM amplitude. Multiple run length drops can be observed here, which would increase the workload of the model

becoming non-informative (with smaller values). We evaluate the model's sensitivity toward α_0, β_0 and provide details in Sect. 4.

In summary, BOCPD first observes and analyzes a continuous stream of data and maintains a reference for typical data patterns as a baseline (the aforementioned stable state). When data significantly deviates from this baseline, the system identifies such a point as a change point. Following this detection, the algorithm adapts its baseline reference to incorporate the new pattern starting from the identified change point. Consequently, the algorithm transitions its baseline to the new data pattern, thereby automating the detection of subsequent change points.

3.2 First order difference

To ensure BOCPD operates effectively across different patients, we normalize the data of each patient with a scalar $s > 0$ which is calculated using the average of low-frequency pure-tone audio-grams at 0.25, 0.5, and 1 kHz, taken from clinical data for each patient from a pre-operative test (within 3 months prior to the surgery). For patients who did not undergo the test, we set the default value to $s = 55$ based on the median of all existing patients. This scalar can represent the residual hearing level of each patient [21], allowing the model to handle individual variability in the baseline hearing level, improving the effectiveness of change point detection.

Although the BOCPD algorithm works well on detecting sudden changes in the data stream, it may still face difficulties considering the trends within our surgical data: A rise in CM at an early stage of CI surgery is expected as the electrode is moving closer to the generator. Figure 3 shows an example of the BOCPD algorithm as directly applied to the ECochG data. The top figure is the original CM amplitude and the bottom one demonstrates the run length returned by BOCPD.

We observe that there are multiple run length drops as data keeps shifting up and down, which indicates multiple change points. We determine drop phases whenever the algorithm detects a change, in which case too many unnecessary checks will be done.

To address this issue, we de-trend the data using the first order difference of the scaled CM amplitude, which is commonly used to remove trends in time series data [22, 23]:

$$dif_1 = 0, dif_t = \frac{x_t - x_{t-1}}{s} \text{ when } t > 1 \tag{11}$$

With this transformation, we now have data that mostly follows a normal distribution centered at $\mu = 0$ with a certain variance (calculated based on a hypothesis test [24]). These properties will change only when the CM amplitude is experiencing a dramatic change. This allows us to omit the cumulative changes and focus on the immediate ones. Figure 4 depicts the underlying idea: as we take the difference of the example data (top) the algorithm can now efficiently capture dramatic changes as expected (bottom).

3.3 Scoring traumatic drops

As we have successfully identified the drop phases, we can now check for traumatic ones. The simplest way to do this

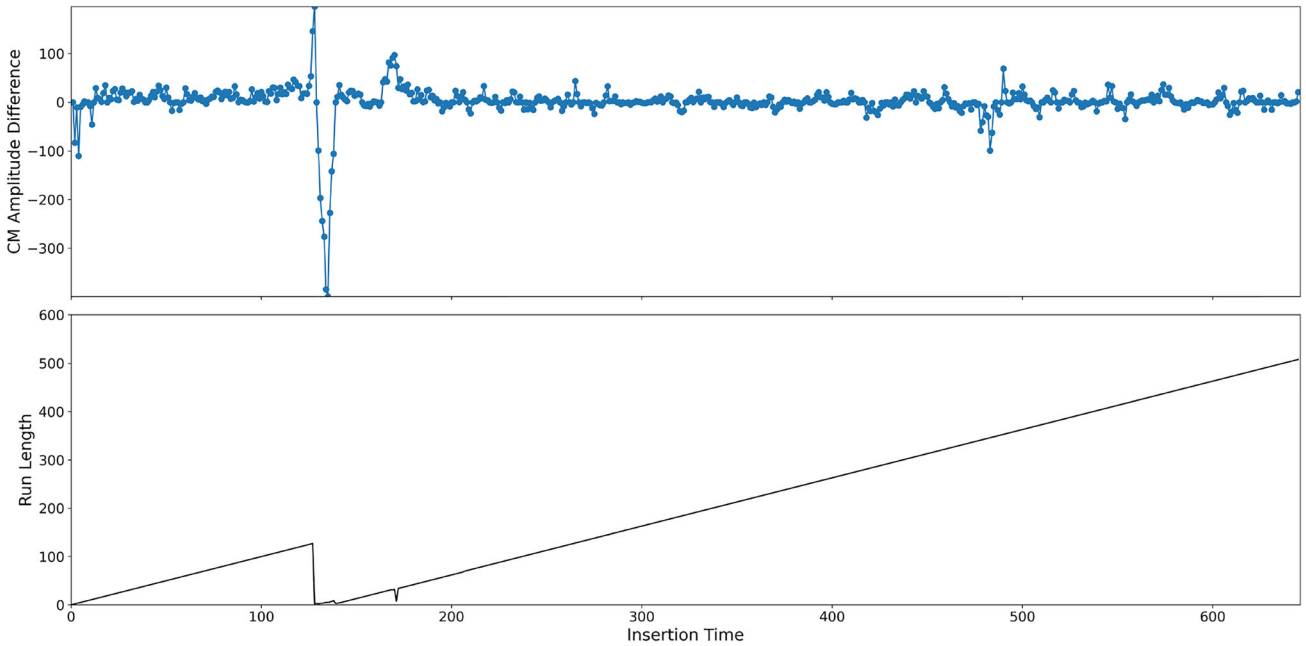


Fig. 4 BOCPD with CM amplitude difference. The run length only drops at a significant change, which reduces the workload of the model subsequently

is to make a decision directly based on the 30% drop rule. However, this approach lacks interpretability as it can only provide a binary decision; a probability score at each point will provide more information to make the decision more explainable. Fixed threshold also lacks flexibility, as it may be beneficial to provide an alert before the 30% rule is satisfied in significant cases, if possible.

To fulfill these desiderata, we develop a scoring function Scr to calculate the probability score for each point. The basis of Scr is the Logistic function $L : \mathbb{R} \rightarrow (0, 1)$, which maps any real number to a value between 0 and 1:

$$L = \frac{1}{1 + e^{-k(x-m)}} \tag{12}$$

where k is the steepness of the curve and m is the sigmoid midpoint on the x-axis. When $x > m$, the Logistic function L is larger than 0.5, suggesting that the corresponding event is likely to occur. Conversely, L falls below 0.5 when $x < m$, indicating a low probability for the corresponding event.

In our task, we define the input x as the peak-value ratio of the current point $x_t / peak_t$, and midpoint $m = 0.7$ to align with the threshold of a 30% drop. Subsequently, whenever the peak-value ratio exceeds the 0.7 threshold, the Logistic function returns a probability greater than 0.5, indicating a significant drop. We keep k as a temperature parameter to increase the model's sensitivity to positive cases:

$$S_1 = \frac{1}{1 + e^{-k(x_t/peak_t - 0.7)}} \tag{13}$$

where $peak_t$ is the peak value for the current drop phase. To determine the k value, we randomly select 10 patients' data and grid-search the optimal value that gives the best performance ($k = 50$).

Based on expert advice, a larger peak (indicating a high point in the CM amplitude) and/or a more significant drop (indicating a sharp decline from the peak) are associated with a higher probability of trauma. To reflect this relationship, we adjust the score to account for peak and difference values and the final formula of the scoring function is written as:

$$Scr = \frac{1}{1 + e^{-k(x_t/peak_t - 0.7)}} \times \frac{peak_t dif_t}{max(\boldsymbol{x}_{1:t})} \tag{14}$$

where $\max(\boldsymbol{x}_{1:t})$ is the maximum observation, used to normalize the peak value.

For those points that exceed the threshold, Scr will assign scores larger than 1 due to the rescaling. We simply recognize them as valid probabilities with a value of 1. All the points with scores larger than 0.5 are detected as traumatic.

Applying the BOCPD algorithm on the first order difference, we can model the data accurately and detect whenever a data property (mean and/or variance) changes. Then, by calculating the difference, we can determine whether the data is rising or dropping. Finally, with the proposed scoring function, we can identify the traumatic points inside drop phases. The overall workflow of BLODD is shown in Algorithm 1.

Algorithm 1 BLODD algorithm

1: **Initialization**:

$$P(r_0 = 0) = 1$$
$$\mu_0 = 0$$
$$\kappa_0 = 1$$
$$\alpha_0 = 0.1$$
$$\beta_0 = 0.1$$

2: **Observe New Data** x_t $(t > 0)$

3: **Calculate posterior**:

$$P(x_t|r_{t-1}, \boldsymbol{x}_t^{(r)}) = t_{2\alpha}\left(\tilde{\boldsymbol{x}}|\mu_{t+1}, \frac{\beta_{t+1}(\kappa_{t+1}+1)}{\alpha_{t+1}\kappa_{t+1}}\right)$$

4: **Calculate Run Length Change Probability**:

$$P(r_t = 0, \boldsymbol{x}_{1:t})$$
$$= \sum_{r_{t-1}} P(r_{t-1}, \boldsymbol{x}_{1:t-1}) P(x_t|r_{t-1}, \boldsymbol{x}_t^{(r)}) H(r_{t-1})$$
$$P(r_t = r_{t-1} + 1, \boldsymbol{x}_{1:t})$$
$$= P(r_{t-1}, \boldsymbol{x}_{1:t-1}) P(x_t|r_{t-1}, \boldsymbol{x}_t^{(r)})(1 - H(r_{t-1}))$$

5: **Calculate Evidence**:

$$P(\boldsymbol{x}_{1:t}) = \sum_{r_t} P(r_t, \boldsymbol{x}_{1:t})$$

6: **Detect Change Point based on Run Length Distribution**:

$$P(r_t|\boldsymbol{x}_{1:t}) = \frac{P(r_t, \boldsymbol{x}_{1:t})}{P(\boldsymbol{x}_{1:t})}$$

7: **Determine Drop Phases via** dif_t

8: **Identify Traumatic Drop Points**:

$$Scr = \frac{1}{1 + e^{k(x_t/peak_t - 0.7)}} \times \frac{peak_t dif_t}{max(\boldsymbol{x}_{1:t})}$$

9: **Update Parameters**:

$$\mu_t = \frac{\mu_{t-1}\kappa_{t-1} + x_t}{\kappa_{t-1} + 1}$$
$$\kappa_t = \kappa_{t-1} + 1$$
$$\alpha_t = \alpha_{t-1} + \frac{1}{2}$$
$$\beta_t = \beta_{t-1} + \frac{\kappa_{t-1}(x_t - \mu_{t-1})^2}{2(\kappa_{t-1} + 1)}$$

10: **Repeat from Step 2**.

4 Experiments

4.1 Evaluation metrics

Traditional metrics used to evaluate a machine learning model are **accuracy**, **recall**, and **precision**. They can reflect how accurately the model performs, how sensitive the model is toward positive cases, and how reliable it is in returning positive predictions, respectively. In addition to these single metrics, the area under the curves (AUC) has also been used to integrate different metrics. The precision-recall curve (PRC), as an example, gives the trade-off between recall and precision with varying thresholds. This makes it prioritize the positive cases which are the primary concern in applications such as surgery. On the other hand, it is also robust to the imbalanced data which is a common issue in surgical datasets [25]. These characteristics make the **area under the precision-recall curve (AUPRC)** a suitable metric for our task.

Beyond these traditional machine learning metrics, it is also encouraged to develop more interpretable metrics for surgical practitioners to capture the nuances of the model. First of all, instead of presenting percentages that are calculated point-wise, we propose a phase-wise metric that directly shows the number of traumatic phases that have been (1) successfully detected, (2) missed, and (3) the number of false alarms. These numbers can provide the surgeon with a straightforward understanding of what they would encounter during the CI surgery with these models. The start of a drop is more important than its end as any remedial action is taken as soon as traumatic drop is detected. Thus, we design a criteria that is based mostly on start points and use end points to separate false alarms from correct detection:

- **Success**: start point of the predicted phase is no more than 5 seconds earlier and 2 seconds later when compared to the actual start, and the end point is later than the actual start.
- **Missed**: the actual phase is not detected by the model or the start point is more than 2 seconds later.
- **Extra**: predicted phases' start and end are outside that of their actual counterparts, or the start point is more than 5 seconds earlier. These phases are regarded as false alarms (false positive), which should be avoided during a real-world detection process.

Here, the actual phases are annotated by domain experts, and the time intervals (5 and 2 s) are also selected based on their advice. For fair comparison, we generate phases from all models ([10, 15, 17], and ours) in the same way. We first join consecutive positive predictions into phases. Then, we combine phases that have a gap less or equal to 5 seconds (e.g., $[1, 4], [7, 10] \rightarrow [1, 10]$) and remove the single-point predictions to reduce sparsity. In addition, to evaluate the model's overall performance on accurately identifying phases while minimizing both missed and extra detections, we report the F_1 score, which is a measure that balances these aspects in a single metric. It is calculated as:

$$F_1 = \frac{2 \cdot Success}{2 \cdot Success + Missed + Extra}.$$

Table 1 Performance change with different parameter settings

Gamma	Accuracy(↑)	Recall(↑)	Precision(↑)	AUPRC(↑)
$Ga(0.01, 0.01)$	0.975	0.805	0.839	0.789
$Ga(0.05, 0.05)$	0.975	0.813	0.840	0.803
$Ga(0.1, 0.1)$	0.975	0.813	0.840	0.803
$Ga(0.5, 0.5)$	0.975	0.813	0.840	0.803
$Ga(1, 1)$	0.975	0.805	0.839	0.795

↑ represents higher values are better

Table 2 Point-wise machine learning metrics

Models	Accuracy(↑)	Recall(↑)	Precision(↑)	AUPRC(↑)
AdaBoost Tree	0.956	**0.861**	0.436	0.665
Gaussian Process	0.827	0.776	0.451	0.487
RuLSIF	0.946	0.858	0.602	0.594
BLODD (ours)	**0.971**	0.819	**0.801**	**0.778**

↑ represents higher values are better. Best performances are highlighted

Table 3 Phase-wise surgical metrics

Models	Success(↑)	Missed(↓)	Extra(↓)	$\mathbf{F_1}$(↑)
AdaBoost Tree	**119**	**4**	146 (1.87/p)	0.613
Gaussian Process	116	7	139 (1.78/p)	0.614
RuLSIF	118	5	64 (0.82/p)	0.774
BLODD (ours)	118	5	**18 (0.23/p)**	**0.911**

↓ represents lower values are better. "/p" represents per-patient. Best performances are highlighted

4.2 Parameter sensitivity

We use the data from the same 10 patients that was used to perform the grid-search for k to check how sensitive the model is to different values of α_0 and β_0. We record the performances of the model with respect to the previously introduced machine learning metrics under different α_0 and β_0 values. Table 1 lists the results, from which we can conclude the model is not very sensitive to these parameters as the performance is mostly stable. Additionally, the fact that the results around $Ga(0.1, 0.1)$ are slightly better than that at higher parameter values ($Ga(1, 1)$) and lower parameter values $Ga(0.01, 0.01)$ aligns with the findings in [20], which suggests our choice for the initial values was appropriate.

4.3 Evaluation results

We first show the evaluation results on 78 patients' data from an older system of ECochG (excluding the patients we used for hyperparameter tuning). The traditional machine learning metrics are listed in Table 2 and the newly proposed surgical metrics are presented in Table 3.

In previous work [10], the model (AdaBoost Tree) was selected based on sensitivity, aiming at identifying as much traumatic drops as possible. As such, their model achieved a higher recall score. However, this makes their model over-sensitive to the traumatic points and generates too many false alarms. On the other hand, although we also increase the model's sensitivity by adjusting the temperature, we manage to mitigate the over-sensitivity issue by rescaling the score. According to Table 3, the proposed BLODD misses only one more traumatic drop than the tree method while reducing the number of false alarms by nearly 90%. The **AUPRC** in Table 2 also suggests our model achieves a better equilibrium between recall and precision. The average prediction time (per time-point) of BLODD is $1.2ms$ which is only $0.5ms$ slower than the AdaBoost Tree during the evaluation stage. This is an acceptable prediction time for this application, especially considering that there is no prior training required. When comparing with other OCPD methods, our model also performs better in general, considering all metrics. The BOCPD automatically adapts to the changes in data by maintaining a posterior distribution over the run length. This dynamic adaptability allows the model to effectively detect shifts even in non-stationary and noisy data, where the other two OCPD methods often struggle or fail. This justifies our choice of BOCPD for the drop phase detection task. Overall, these results show that BLODD is a more reliable method for the task of real-time drop detection in CI surgery.

Table 4 Point-wise machine learning results for new patients

Models	Accuracy(↑)	Recall(↑)	Precision(↑)	AUPRC(↑)
AdaBoost Tree	0.867	0.771	0.459	0.393
BLODD (ours)	**0.945**	**0.823**	**0.818**	**0.766**

↑ represents higher values are better. Best performances are highlighted

Table 5 Phase-wise surgical results for new patients

Models	Success(↑)	Missed(↓)	Extra(↓)	$\mathbf{F_1}$(↑)
AdaBoost Tree	17	2	8 (1.6/p)	0.773
BLODD (ours)	**19**	**0**	**3 (0.6/p)**	**0.927**

↓ represents lower values are better. "/p" represents per-patient. Best performances are highlighted

4.4 Generalization

An important characteristic that is expected for a surgical detection model is generalizability. As the patients' characteristics vary greatly from each other, the model should be able to work well not only on existing patients, but also on new ones. To evaluate the models' generalizability, we collected 5 new patients' data from a different ECochG system to that used in the previous experiments. This new data contains more noise, including impulse noise which would fundamentally impact the detection. We remove this impulse noise using a median filter (with a window size of 5, which was determined by a grid-search) first, to ensure compatibility with previous data. This pre-processing is implemented in real time. The data are extended by reflecting the edge values ($cba|abc|cba$) to avoid errors at the early stage (where the number of observations are less than the window size). We only compare the proposed method and the previous work [10] here. Tables 4 and 5 list the results from both models when run on the new data.

Our proposed BLODD shows consistent performance compared to the results of the data obtained from the previous ECochG system, while the tree method experiences a performance drop: their recall score is even lower than ours now. Figure 5 below provides the visual results of these patients to help with assessment. We mark the traumatic points predicted by the model with $\times$ and the true traumatic drop phases with green bars.

The good results of our model on the newly collected data reflect the generalizability of the proposed model, and indicates the potential for its future deployment to assist in CI surgery.

5 Limitations and future perspectives

Although our model demonstrates robust performance with respect to various metrics, we acknowledge its limitations and propose future work aimed at mitigating these issues and preparing it for real-world deployment. Firstly, the current algorithm is designed for research purposes only. More comprehensive testing and validation from multiple clinical sites, especially using data collected from the relevant commercially-available ECochG systems, is required. Secondly, This work includes only the CM amplitude in the prediction of traumatic drops and subsequent cochlear trauma, as the work is developed upon the existing trauma detection & reversal approach [7, 9]. Current literature is debating the utility of including additional ECochG components for trauma detection, including the latency of the CM, as well as the ANN [10, 26]. The inclusion of these components in additional machine learning approaches, in order to optimize the components that best predict hearing preservation, is a goal of future work. Thirdly, the current model is primarily algorithmic, whose operation might be difficult to understand for practitioners or clinicians not specialized in machine learning. An interface that conveys useful information and functionality (such as possibility scores and trauma alerts) to the practitioners in an easily understandable way should be implemented. Specifically, intuitive visual elements (e.g., confidence bars) are expected to be implemented to indicate the confidence levels of the model's predictions. Such visual cues will allow clinicians to quickly assess the certainty associated with each output at a glance. Additionally, visual highlights (e.g., ranked bar chart) of the top contributing risk factors can be included within the interface to bridge the gap between the algorithm outputs and clinical decision-making. Finally, practitioners should be trained on how to properly interpret the outputs of the model, so that remedial action can be undertaken immediately.

The current model does not indicate specific interventions or clinical advice, as the scope of the model is rapid and accurate drop detection. The development of this algorithm into a clinically viable tool, i.e., one that indicates specific recommendations to the clinician during cochlear implantation, is part of future work that should be undertaken in collaboration with a commercial cochlear implant manufacturer. At present, basic feedback to the surgeon on the current condition of the CM has been demonstrated to be usable to significantly improve hearing preservation [9], suggesting that an in-depth understanding of the approach of the tool is unlikely to be necessary for the approach to be clinically viable.

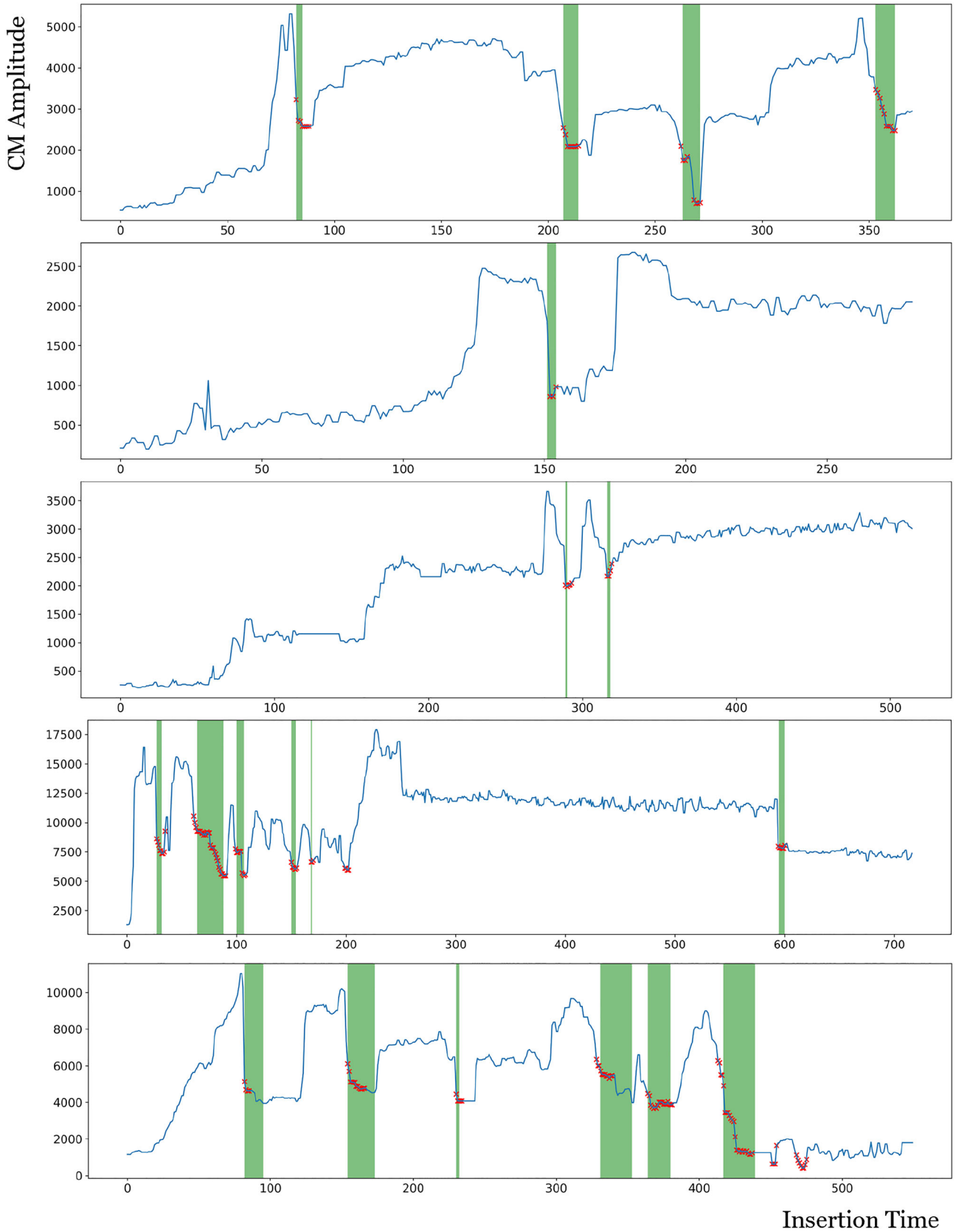


Fig. 5 Visualization results on new patients data, × are predicted traumatic points and green bars are expert assessment

Regarding ethical and safety concerns, cochlear implant surgeons are trained to use a series of real-time health monitors during implantation, with the closest example to this system being a nerve function monitor during drilling. The real-time feedback proposed to be used with this algorithm works seamlessly with the "slow and careful" implantation approach that is recommended to all implant surgeons due to the current generation of highly-flexible electrode arrays. Surgeons are, as always, directed to use their clinical expertise in how to interpret feedback from the system. For example, if cochlear trauma is detected very early in implantation, which is a relatively rare occurrence, the surgeon will make the decision on whether to advance the array to complete insertion. This complete insertion will result in all electrodes being within the cochlea, a highly desirable insertion metric. However it is unlikely to result in optimal hearing preservation. The surgeons are already trained to make trade-off decisions of this nature through implantation. At present, it is expected around 70% of patients who receive a cochlear implant will lose normal function in that ear. However, the electrical hearing provided by the implant is sufficiently superior that this is an acceptable trade-off.

6 Conclusion

In this paper, we developed an efficient and reliable model for drop detection during cochlear implant (CI) surgery based on well-established methods. We first applied the Bayesian method to find drop phases in real time and then formulated a scoring function, derived from the logistic function, to compute probability scores for each point inside the drop phases to detect traumatic drops. The proposed method (BLODD) is an unsupervised and purely data-driven model, ensuring its adaptability across varying patient data. The empirical evaluations demonstrated the reliability of the BLODD as it significantly reduces the false alarms while keeping the sensitivity at a high level. The detection time is within requirements for real-time drop detection. The consistent performance of BLODD on new data further suggests its generalizability to new patients. These aspects indicate the effectiveness and efficiency of the BLODD for drop detection and highlight its potential to be used in real-world CI surgery.

Acknowledgements This work is partially supported by NHMRC Development Grant Project 2000173, The University of Melbourne's Research Computing Services and the Petascale Campus Initiative.

Author contributions Y.L. contributed to conceptualization, methodology, data analysis, writing—original draft, writing—review and editing S.W. contributed to supervision, writing—review and editing C.B. contributed to data collection, writing—review and editing S.O. contributed to data collection, writing–review and editing J.B. contributed to supervision, funding acquisition, writing—review and editing.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Data Availability The data underlying the results presented in the study are available at https://doi.org/10.5061/dryad.931zcrjmz.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Haile, L.M., Kamenov, K., Briant, P.S., Orji, A.U., Steinmetz, J.D., Abdoli, A., Abdollahi, M., Abu-Gharbieh, E., Afshin, A., Ahmed, H., et al.: Hearing loss prevalence and years lived with disability, 1990–2019: findings from the global burden of disease study 2019. Lancet **397**(10278), 996–1009 (2021)
2. Gifford, R.H., Dorman, M.F., Shallop, J.K., Sydlowski, S.A.: Evidence for the expansion of adult cochlear implant candidacy. Ear Hear. **31**(2), 186 (2010)
3. Eshraghi, A.A., Nazarian, R., Telischi, F.F., Rajguru, S.M., Truy, E., Gupta, C.: The cochlear implant: historical aspects and future prospects. Anat. Rec. Adv. Integr. Anat. Evol. Biol. **295**(11), 1967–1980 (2012)
4. Carlson, M.L., Driscoll, C.L., Gifford, R.H., Service, G.J., Tombers, N.M., Hughes-Borst, B.J., Neff, B.A., Beatty, C.W: Implications of minimizing trauma during conventional cochlear implantation. Otol. Neurotol. Offl. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol. **32**(6), 962 (2011)
5. Holden, L.K., Finley, C.C., Firszt, J.B., Holden, T.A., Brenner, C., Potts, L.G., Gotter, B.D., Vanderhoof, S.S., Mispagel, K., Heydebrand, G., et al.: Factors affecting open-set word recognition in adults with cochlear implants. Ear Hear. **34**(3), 342 (2013)
6. Campbell, L., Kaicer (Umansky), A., Sly, D., Iseli, C., Wei, B., Briggs, R., O'Leary, S.: Intraoperative real-time cochlear response telemetry predicts hearing preservation in cochlear implantation. Otol. Neurotol. **37**, 1 (2016)
7. O'Leary, S., Briggs, R., Gerard, J.-M., Iseli, C., Wei, B.P., Tari, S., Rousset, A., Bester, C.: Intraoperative observational real-time electrocochleography as a predictor of hearing loss after cochlear implantation: 3 and 12 month outcomes. Otol. Neurotol. **41**(9), 1222 (2020)
8. Weder, S., Bester, C., Collins, A., Shaul, C., Briggs, R.J., O'Leary, S.: Toward a better understanding of electrocochleography: analysis of real-time recordings. Ear Hear. **41**(6), 1560–1567 (2020)
9. Bester, C., Collins, A., Razmovski, T., Weder, S., Briggs, R.J., Wei, B., Zakaria, A.F., Gerard, J.-M., Mitchell-Innes, A., Tykocinski, M., et al.: Electrocochleography triggered intervention successfully

preserves residual hearing during cochlear implantation: Results of a randomised clinical trial. Hear. Res. **426**, 108353 (2022)
10. Wijewickrema, S., Bester, C., Gerard, J.-M., Collins, A., O'Leary, S.: Automatic analysis of cochlear response using electrocochleography signals during cochlear implant surgery. Plos one **17**(7), e0269187 (2022)
11. Dalbert, A., Huber, A., Veraguth, D., Roosli, C., Pfiffner, F.: Assessment of cochlear trauma during cochlear implantation using electrocochleography and cone beam computed tomography. Otol. Neurotol. **37**(5), 446–453 (2016)
12. Basseville, M., Nikiforov, I.V.: Detection of Abrupt Changes: Theory and Application. Prentice-Hall Information and System Sciences Series, vol. 104. Prentice Hall, Hoboken (1993)
13. Kawahara, Y., Sugiyama, M.: Sequential change-point detection based on direct density-ratio estimation. Stat. Anal. Data Min. ASA Data Sci. J. **5**(2), 114–127 (2012). https://doi.org/10.1002/sam.10124
14. Liu, S., Yamada, M., Collier, N., Sugiyama, M.: Change-point detection in time-series data by relative density-ratio estimation. Neural Netw. **43**, 72–83 (2013). https://doi.org/10.1016/j.neunet.2013.01.012
15. Yamada, M., Kimura, A., Naya, F., Sawada, H.: Change-point detection with feature selection in high-dimensional time-series data. In: IJCAI International Joint Conference on Artificial Intelligence (2013)
16. Adams, R.P., MacKay, D.J.C.: Bayesian Online Changepoint Detection (2007)
17. Garnett, R., Osborne, M.A., Roberts, S.J.: Sequential bayesian prediction in the presence of changepoints. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 345–352 (2009)
18. Puza, B.: Bayesian Methods for Statistical Analysis. ANU Press, Canberra (2015)
19. Murphy, K.P.: Conjugate Bayesian analysis of the Gaussian distribution. def **1**, 16 (2007)
20. Lemoine, N.P.: Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. Oikos **128**(7), 912–928 (2019)
21. Andonie, R.R., Wimmer, W., Schraivogel, S., Mantokoudis, G., Caversaccio, M., Weder, S.: Electrocochleography in cochlear implant recipients: correlating maximum response with residual hearing. Ear Hear. 10–1097 (2024)
22. Wen, Q., Gao, J., Song, X., Sun, L., Tan, J.: Robusttrend: A huber loss with a combined first and second order difference regularization for time series trend filtering. arXiv preprint arXiv:1906.03751 (2019)
23. Mills, T.C.: Unit roots, difference and trend stationarity, and fractional differencing. In: Mills, T.C. (ed.) Applied Time Series Analysis, pp. 71–101. Academic Press, USA (2019). https://doi.org/10.1016/B978-0-12-813117-6.00005-3
24. D'Agostino, R., Pearson, E.S.: Tests for departure from normality. Empirical results for the distributions of $b2$ and $\sqrt{b1}$. Biometrika **60**(3), 613–622 (1973)
25. Saito, T., Rehmsmeier, M.: The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE **10**(3), 1–21 (2015). https://doi.org/10.1371/journal.pone.0118432
26. Andonie, R.R., Wimmer, W., Wildhaber, R.A., Caversaccio, M., Weder, S.: Real-time feature extraction from electrocochleography with impedance measurements during cochlear implantation using linear state-space models. IEEE Trans. Biomed. Eng. **70**(11), 3137–3146 (2023)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.