Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

SpikeTrainPatternDetector: a provenance-aware R/Shiny framework for neuronal firing-pattern detection and review

Houchun Zhou

zhouhouchun@outlook.com

Semenov Institute of Chemical Physics

Alexey Sedov

Semenov Institute of Chemical Physics

Research Article

Keywords: spike train, firing-pattern detection, neuroinformatics, burst, tonic firing, reproducibility

Posted Date: July 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10115382/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

SpikeTrainPatternDetector: a provenance-aware R/Shiny framework for neuronal firing-pattern detection and review

Houchun Zhou[1,2] · **Alexey Sedov**[1,2]

1 N.N. Semenov Federal Research Center for Chemical Physics, Russian Academy of Sciences, Moscow, Russia
2 Moscow Center for Advanced Studies, Moscow, Russia
Corresponding author: Houchun Zhou (zhouhouchun@outlook.com)
ORCID: 0009-0009-0575-8279
Running title: Provenance-aware firing-pattern review
Article type: Software Original Article
Keywords: spike train; firing-pattern detection; neuroinformatics; burst; tonic firing; reproducibility

Abstract

Operational labels such as bursts, pauses, tonic firing and high-frequency activity depend on decisions that are often lost after analysis, including timestamp exclusions, threshold selection, boundary rules and manual revisions. We developed SpikeTrainPatternDetector (STPD), an open-source R/Shiny package that preserves these decisions as part of the analysis record. STPD recomputes inter-spike intervals (ISIs) from timestamps, performs quality control (QC) before detection, records threshold values and their sources, generates candidate events, resolves overlapping candidates deterministically, protects manual annotations and exports public event tables together with diagnostic and reproducibility records. Package execution was checked with a bundled timestamp fixture containing 23 spike trains and 16,728 timestamps. Detector behavior was then assessed on 30 simulated 30-s spike trains comprising 8,746 labeled ISIs. With automatically resolved thresholds, six-class interval-level scoring yielded accuracy = 0.703, macro-F1 = 0.498 and weighted F1 = 0.669. High-frequency spiking was recovered most reliably (F1 = 0.906), whereas burst-family, tonic, pause and high-frequency tonic labels were less stable. Simulator-informed thresholds increased weighted F1 to 0.712 but did not resolve competition between high-frequency states. Mean-ISI and LogISI support methods overcalled burst intervals on the same benchmark. STPD is therefore intended as provenance-preserving infrastructure for candidate generation, expert review and preparation-specific validation, rather than as a preparation-independent biological classifier.

Introduction

Neuronal activity is recorded as a sequence of action-potential times, but many experimental questions concern structured episodes within that sequence. Bursts, pauses, tonic discharge and sustained high-frequency activity are used to describe changes in synaptic drive, network state, stimulation response, pathology and behavior. These patterns differ in the temporal evidence needed for annotation: burst definitions emphasize compact groups of short ISIs and their flanking intervals, pause definitions emphasize isolated long gaps, and tonic or high-frequency states depend on duration, rate and regularity. Several such regimes can occur within a single recording (Fig. 1).

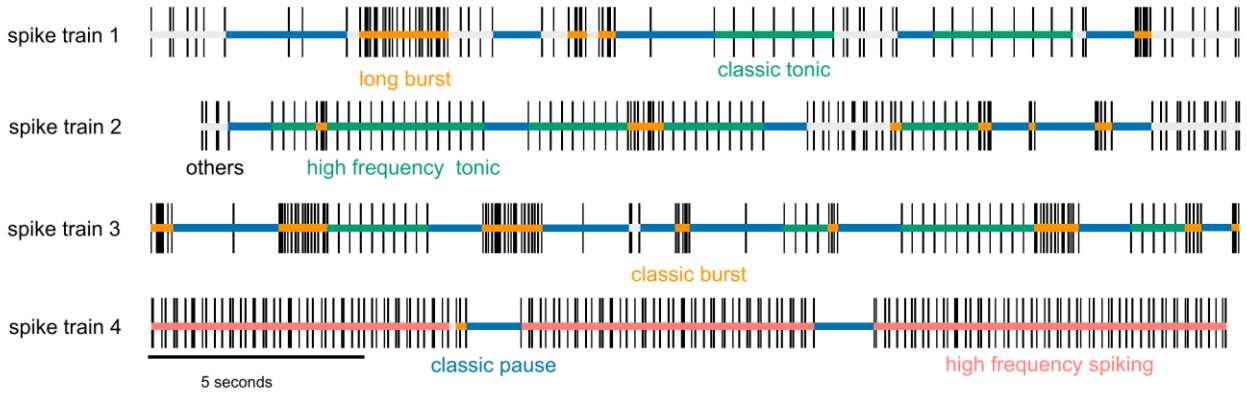


Fig. 1 *Simulated spike trains illustrating the annotation problem. Vertical ticks indicate spike times; colored intervals show known simulated regimes, including burst-family events, pauses, tonic firing, high-frequency tonic activity and high-frequency spiking. The figure is schematic; quantitative performance was evaluated separately against withheld simulator labels.*

These labels are operational rather than universal. Thresholds for spike count, ISI duration, local contrast, firing rate and regularity vary across preparations and analytical aims. The same timestamp sequence can also produce different annotations

when duplicate spikes, non-monotonic ordering, refractory-period violations, nonstationarity or manual corrections are handled differently. A final label is therefore difficult to reproduce when the data exclusions, parameter sources and boundary decisions that produced it are not retained.

The annotation problem is compounded by the way label vocabularies mix distinct biological and analytical dimensions. A burst or pause is a transient event; tonic firing is a sustained state; high frequency is a rate attribute; possible_burst denotes review status; and unclassified denotes abstention rather than a physiological regime. Software may nevertheless require a single label track for display and export. When such categories compete for the same interval, the arbitration rule becomes part of the scientific result and should be visible to the user.

Existing methods address parts of this problem. The coefficient of variation and local variation summarize global and local ISI variability (Holt et al., 1996; Shinomoto et al., 2005). Mean-ISI and LogISI methods provide adaptive evidence for burst-related short-interval structure (Chen et al., 2009; Pasquale et al., 2010). Broader software ecosystems support electrophysiology data models, spike-train analysis, synchrony measures or spike-sorting workflows. These methods are valuable, but they do not by themselves preserve the complete path from imported timestamps to a reviewed multi-pattern annotation.

In routine analysis, the missing information is often procedural: whether timestamps were reordered, which ISIs were excluded from threshold estimation, how a threshold was obtained, which competing candidates were rejected, and whether a reviewer changed the automatic call. Retaining only the final labels prevents independent inspection of these decisions and makes preparation-specific validation difficult.

SpikeTrainPatternDetector (STPD) was developed to make this decision path explicit. The package combines QC-first timestamp handling, configurable candidate rules, deterministic interval selection, manual-label protection and exportable provenance records in an R/Shiny workflow. Literature-linked Mean-ISI and LogISI procedures are retained as support layers rather than merged silently into the main event grammar. The resulting architecture separates candidate generation, provenance capture, reference-based validation, and downstream visualization (Fig. 2).

This study has three aims: to describe the software architecture and the records generated at each stage; to demonstrate reproducible execution on a bundled timestamp fixture; and to use a controlled synthetic benchmark to identify both successful and unstable operating regimes. The benchmark evaluates the current release as implemented. It does not treat simulator-informed thresholds or automatic labels as estimates of universal biological truth.

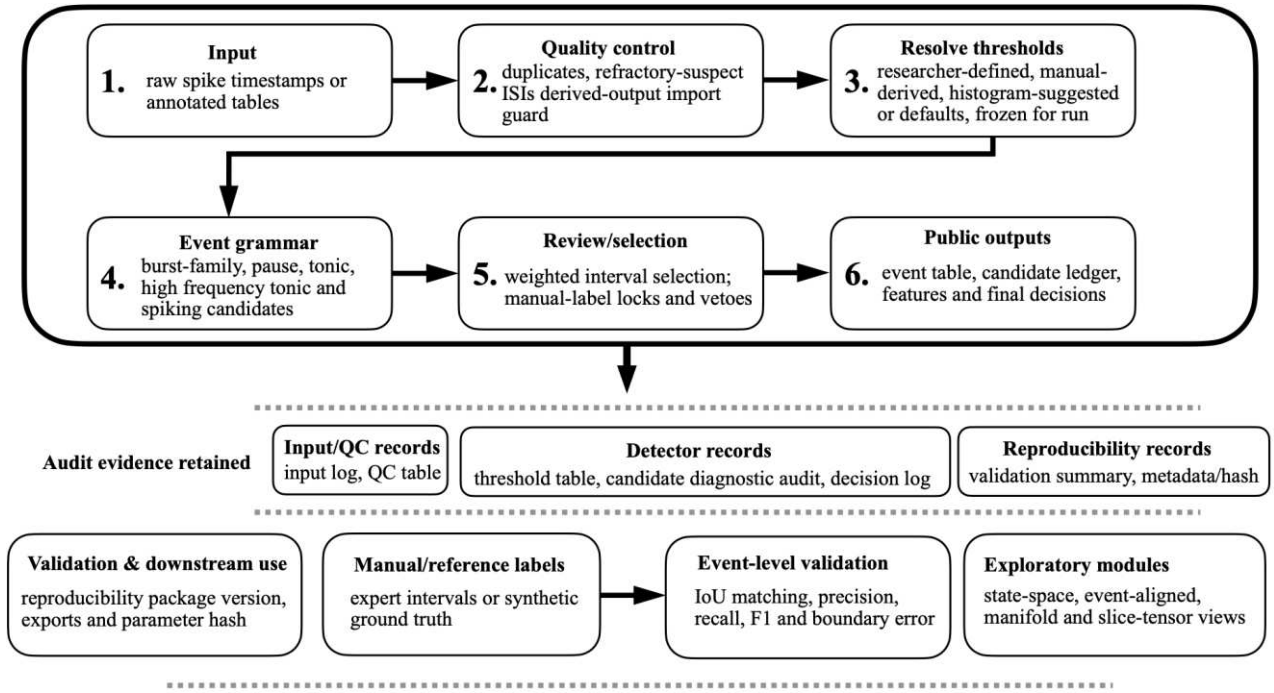


Fig. 2 STPD workflow from spike timestamps to reviewable event calls. *Raw or annotated spike-timestamp tables pass through six stages: import, pre-detection quality control, threshold resolution, event-grammar candidate generation, weighted interval selection with manual-label locks or vetoes, and export of event tables, candidate ledgers, features, and final decisions. Throughout the run, STPD retains input/QC records, detector records, and reproducibility records, including threshold sources, candidate diagnostics, decision logs, software version, metadata, and parameter hashes. When expert annotations or synthetic ground truth are available, event-level agreement is evaluated using intersection-over-union matching, precision, recall, F1, and boundary error. Exploratory state-space, event-aligned, manifold, and slice-tensor views are downstream modules and do not contribute hidden evidence to the final labels.*

Methods

Software architecture and provenance model

STPD version 1.2.1 is implemented primarily in R and provides an interactive interface built with Shiny. Plotly is used for spike-train and diagnostic views. Native C routines accelerate repeated operations, including ISI-percentile calculation, cached local-median estimation, short-ISI run scanning, structural candidate scanning and interval-overlap calculation; these routines do not introduce a separate classification model.

The package exposes stable entry points for dataset construction, detector execution, export and reference-label comparison. Parameters are materialized from a YAML-backed contract that defines defaults, validation rules, user-interface metadata and the public spiketrainpattern namespace. The same contract is used by programmatic and graphical workflows, reducing discrepancies between interface settings and scripted analyses.

The workflow separates four records that are often conflated. The input record documents timestamp integrity and import decisions. The parameter record stores effective thresholds together with their sources. The detector record retains proposed, rejected, suppressed and selected candidates. The review record stores the relationship between automatic evidence, manual labels and final exported calls. A run-level record links these outputs to the software version, input manifest and parameter hash.

This design does not assume that a rule-based call is a biological ground truth. Automatic results are treated as inspectable evidence that can be screened, rejected or compared with independent labels. Performance claims are therefore restricted to datasets for which withheld simulator labels, expert annotations or another independent endpoint are available.

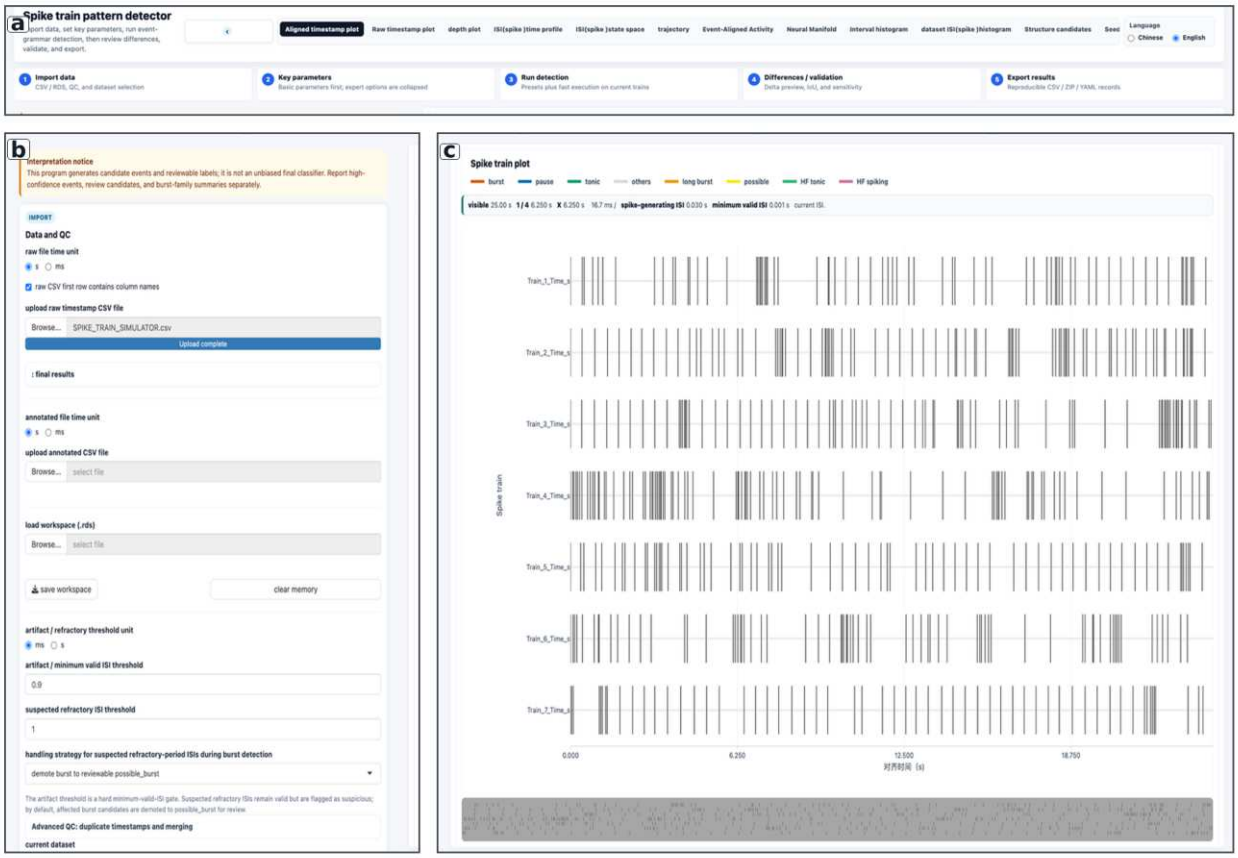


Fig. 3 *Graphical interface. (A) Import, quality-control and parameter controls. (B) Aligned spike-train review with the staged workflow for detection, comparison and export. (C)Spike train raster plot. The displayed data are a software demonstration, not a clinical inference.*

Installation and minimal execution

STPD is distributed as an R package through a versioned public repository and archived release. A minimal scripted analysis imports timestamp columns, materializes the parameter contract, runs detection and writes the public and diagnostic outputs:

```
params <- default_params()
ds <- build_spike_dataset("spikes.csv", mode = "raw", unit_in = "s")
ds <- stpd_detect(ds, params, lock_manual = TRUE, collect_diagnostics = TRUE)
stpd_export_results(ds, params, out_dir = "results")
```

The Shiny application exposes the same import, QC, threshold inspection, candidate review, reference-label comparison and export sequence. A formal report should identify the package version or commit, R version, operating system, dependency versions, input and parameter hashes, QC policy, resolved thresholds, label crosswalk and exported event and audit files.

Table 1 Software metadata and minimum reproducibility record

Field	Current value or reporting requirement
Software	SpikeTrainPatternDetector (STPD)
Purpose	Provenance-aware candidate generation, review and validation of neuronal firing-pattern annotations
Implementation	R package with Shiny interface; native C routines for repeated ISI and interval operations
Release	Version 1.2.1
License	MIT + file LICENSE
Repository	https://github.com/zhouhouchun/SpikeTrainPatternDetector-STPD
Archived release	https://doi.org/10.5281/zenodo.20650295
Core outputs	QC table, threshold table, candidate ledger, diagnostic candidate audit, event table, decision audit and run metadata

Field	Current value or reporting requirement
Minimum record	Software/commit version, R/OS/dependencies, input hash, parameter hash, QC policy, threshold sources and scoring crosswalk

Input data model

The primary input is a table of spike timestamps. In raw CSV mode, each non-empty numeric column is interpreted as one spike train. Timestamps may be supplied in seconds or milliseconds and are converted internally to seconds. Missing values are removed, the original order is checked, timestamps are sorted within train and all ISIs are recomputed from adjacent timestamps.

For an ordered spike train

$$t_1 < t_2 < \cdots < t_n,$$

the inter-spike intervals are

$$\Delta_i = t_i - t_{i-1}, \qquad i = 2, \ldots, n,$$

with Δ1 undefined. A candidate is represented as a contiguous ISI-indexed interval

$$c = [a, b], \qquad 2 \le a \le b \le n.$$

Candidate c contains the ISIs

$$D(c) = \{\Delta_a, \Delta_{a+1}, \ldots, \Delta_b\}.$$

Because Δa lies between ta-1 and ta, the corresponding candidate spans one more spike than ISI. Therefore,

$$N_{\mathrm{ISI}}(c) = b - a + 1,$$

$$N_{\mathrm{spk}}(c) = b - a + 2,$$

$$t_{\mathrm{start}}(c) = t_{a-1}, \qquad t_{\mathrm{end}}(c) = t_b,$$

and

$$T(c) = t_b - t_{a-1}.$$

This indexing convention keeps candidate evidence aligned with timestamp-defined onset, offset and overlap calculations.

Annotated inputs may include manual or external labels. External ISI columns are retained only for consistency checks; timestamp-derived ISIs remain authoritative for detection. Files or columns that resemble STPD outputs, including candidate ledgers, threshold tables and final-event tables, are blocked during raw import unless the user explicitly overrides the derived-output guard.

Evaluation datasets and scoring design

Quantitative detector performance was evaluated with synthetic spike trains for which every scorable ISI had a withheld simulator label. The detector received timestamp columns only; reference labels were joined to the output after detection. This separation prevents direct use of the scoring table during the automatic run.

The primary benchmark comprised 30 spike trains of 30 s each. STPD and the simulator use different vocabularies, so a scoring crosswalk was specified before metrics were calculated (Table 2). STPD burst, long_burst and possible_burst outputs were pooled as a burst family for this comparison. Simulator Noisy intervals were mapped to the detector's unclassified category as an operational abstention comparison; the mapping does not imply biological equivalence.

Two STPD configurations were examined. The primary run used automatic threshold resolution. A second run used class-specific ISI bands taken from the simulator configuration. The latter is an oracle-style sensitivity analysis: it estimates the effect of favorable threshold information and is not a deployable real-data setting.

Clinical microelectrode recordings from the globus pallidus internus in patients with Parkinson's disease were used only to set plausible firing-rate and ISI ranges for simulation. These recordings were acquired during standard-of-care deep brain stimulation surgery for electrophysiological target localization and electrode-placement guidance; no research-specific electrode trajectory, recording session, stimulation procedure or other invasive intervention was added. The calibration set contained 198 usable units and 279,240 positive ISIs; the median firing rate was 83.38 Hz and the median ISI was 7.89 ms. The 95th and 99th ISI percentiles were 33.36 and 74.65 ms, respectively.

All clinical datasets were coded or de-identified before analysis, and STPD did not receive direct patient identifiers, raw clinical records, raw voltage traces or waveform features. All participants provided informed consent covering the clinical recording, secondary research use, publication of scientific findings and sharing of coded or de-identified data; the corresponding uses were also authorized by the respective data-holding medical institutions. The clinical recordings had no independent expert event labels and therefore contributed no detector-accuracy estimates or claims about deep brain stimulation physiology.

Table 2 Pre-specified crosswalk for interval-level comparison of simulator and STPD labels

Simulator label	n	Scored category	Scoring treatment
Burst	498	burst family	STPD burst, long_burst and possible_burst outputs were pooled; possible_burst denotes review status in the software
high_frequency_spiking	4,381	high-frequency spiking	Direct vocabulary match
high_frequency_tonic	1,537	high-frequency tonic	Direct vocabulary match; missing predictions contribute F1 = 0 to inclusive macro-F1
Pause	149	pause	Direct vocabulary match
Tonic	1,279	tonic	Direct vocabulary match
Noisy	902	unclassified	Operational abstention mapping because STPD has no Noisy output; not a claim of biological equivalence

Pre-detection quality control

QC is performed before burst, pause, tonic or high-frequency evidence is calculated. For each train, STPD reports spike count, duration, firing rate, raw and valid minimum ISIs, exact duplicate timestamps, non-positive timestamp steps, hard-artifact ISIs, refractory-suspect ISIs, timestamp-ISI mismatches and stationarity warnings. The default hard-artifact threshold is 0.0009 s and the default refractory-suspect threshold is 0.001 s. These values are presets, not universal physiological constants, and must be reported or changed for the recording modality under study.

Let θart denote the hard-artifact threshold and θref the refractory-suspect threshold, with θart < θref. A hard-artifact ISI satisfies

$$0 < \Delta_i < \theta_{\text{art}}.$$

A refractory-suspect ISI satisfies

$$\theta_{\text{art}} \leq \Delta_i < \theta_{\text{ref}}.$$

A valid, non-suspect ISI satisfies

$$\Delta_i \geq \theta_{\text{ref}}.$$

Duplicate timestamps and non-positive ISIs are treated as data-integrity errors unless an explicit import policy resolves them. Hard-artifact intervals remain visible in the timestamp-derived train and QC tables but are excluded from distributional evidence and candidate features that require valid ISIs. Refractory-suspect intervals remain analyzable under the default policy but are flagged and can reduce candidate confidence. Stationarity warnings are retained as interpretive risks because global percentiles and pause or tonic thresholds may be unreliable in strongly nonstationary trains.

Operational handling of short and suspect ISIs

STPD preserves short intervals rather than deleting them silently. Intervals below θart are excluded from valid-ISI percentiles, local-median context, threshold summaries and candidate statistics. Their positions and prevalence remain available for inspection.

Intervals between θart and θref are flagged as refractory-suspect. Under the default contract, a burst-family candidate containing such an interval is assigned review-level possible_burst status rather than promoted to a high-confidence burst call. Alternative policies can warn, split, exclude or mark possible multi-unit contamination.

An isolated hard-artifact interval may be treated as a transparent connector only in the final display or label-continuity layer when it is flanked by the same compatible state. This operation does not convert the interval into valid statistical evidence and does not contribute to threshold estimation or candidate acceptance.

Table 3 Operational treatment of timestamp and short-ISI problems

ISI category	Operational rule	Use in statistical evidence	Default handling
Duplicate or non-positive	Duplicate timestamp or $\Delta i \leq 0$	Excluded	Data-integrity error; collapse or report explicitly
Hard artifact	$0 < \Delta i < \theta art$	Excluded from valid-ISI distributions, context and candidate features	Retained in QC; may connect identical flanking labels only in the display layer
Refractory-suspect	$\theta art \leq \Delta i < \theta ref$	Eligible under policies that exclude only hard artifacts, but flagged	Default burst-family status is possible_burst; alternative policies can warn, split or exclude
Valid, non-suspect	$\Delta i \geq \theta ref$	Eligible	Used by the relevant event-family rule

Parameter materialization and threshold resolution

Detector settings are stored in a YAML-backed contract. The public parameter namespace is spiketrainpattern; internal namespaces are derived before execution. Each run exports the effective parameter set, contract-validation results and a parameter hash.

For each event family p and threshold field f, candidate values can originate from researcher-defined settings, manual-label summaries, histogram suggestions or package defaults. Thresholds are resolved once for the selected dataset and then frozen for all selected trains in that run:

$$\theta_{p,f}^{\mathrm{eff}} = \mathrm{first\ finite}_{\pi}\left\{\theta_{p,f}^{\mathrm{researcher}}, \theta_{p,f}^{\mathrm{manual}}, \theta_{p,f}^{\mathrm{histogram}}, \theta_{p,f}^{\mathrm{default}}\right\},$$

where π is the configured source-priority order. The order can favor researcher-defined, manual-derived, histogram-derived or default values. After source selection, family-specific constraints are applied where required, for example

$$0 \leq \theta_{p,\mathrm{lower}}^{\mathrm{eff}} < \theta_{p,\mathrm{upper}}^{\mathrm{eff}} \leq \theta_{p,\mathrm{bridge}}^{\mathrm{eff}}.$$

The exported threshold table retains every candidate value, the selected value and its source. This record allows a reviewer to determine whether a result changed because the data changed, a threshold source changed or the downstream event grammar changed.

In the present benchmark, withheld simulator labels were introduced only after detection for scoring. The clinical calibration recordings had no expert event labels and were not used to estimate detector accuracy.

Table 4 Parameter families, detector roles and required reporting

Family	Representative fields	Detector role	Required report
QC policy	artifact and refractory thresholds; duplicate policy	Defines analyzable ISIs before detection	QC table and parameter report
Burst family	seed band, bridges, flank contrast, spike-count range	Proposes compact short-ISI events and review-level candidates	Effective thresholds and sources
Pause	absolute floor; local/global context	Proposes long-gap events	Floor and contextual evidence
Tonic	mid-ISI band; duration; CV/LV/MM	Proposes sustained regular states	Band and regularity criteria
High frequency	Q80/Q90/Q95; tolerated gaps; burst-packet guard	Proposes sustained high-rate states	State criteria and suppressions
Manual policy	lock_manual; negative veto; validation mode	Separates reviewer decisions from automatic evidence	Manual-lock and validation settings

Event-grammar detector

The core detector is a deterministic rule system over ISI-indexed candidate intervals. For candidate c, STPD calculates the evidence vector

$$E(c) = [Q_{10}(c), Q_{50}(c), Q_{80}(c), Q_{90}(c), Q_{95}(c), CV(c), LV(c), MM(c), B(c), F_B(c), S_{\min}(c), R(c)],$$

where Qq(c) is the qth within-candidate ISI quantile; CV and LV measure global and local variability; MM is the maximum-to-mean ratio; B and FB are bridge count and bridge fraction; Smin is the minimum flank contrast; and R summarizes regularity or state evidence.

For the m valid positive ISIs in candidate c, in temporal order,

$$CV(c) = \frac{\mathrm{sd}[D(c)]}{\mathrm{mean}[D(c)]},$$

$$LV(c) = \frac{3}{m-1}\sum_{j=1}^{m-1}\left(\frac{\delta_{j+1}-\delta_j}{\delta_{j+1}+\delta_j}\right)^2,$$

and

$$MM(c) = \frac{\max[D(c)]}{\mathrm{mean}[D(c)]}.$$

CV and LV require at least two valid positive ISIs, whereas MM requires at least one. Undefined values remain missing and are handled explicitly by the relevant family rule rather than converted into positive evidence.

Burst-family candidates

Burst-family detection begins with compact seed runs. An ISI supports a seed when

$$\theta^{\mathrm{eff}}_{\mathrm{seed,low}} \le \Delta_i \le \theta^{\mathrm{eff}}_{\mathrm{seed,high}}.$$

Consecutive seed-supporting intervals form a run. Runs shorter than the configured minimum are discarded; retained runs can expand through a limited number of bridge intervals satisfying

$$\theta^{\mathrm{eff}}_{\mathrm{seed,high}} < \Delta_i \le \theta^{\mathrm{eff}}_{\mathrm{bridge}}.$$

Expansion is limited by bridge count, bridge fraction and maximum expansion steps. Bridge fraction is

$$F_B(c) = \frac{B(c)}{N_{\mathrm{ISI}}(c)}.$$

Candidate compactness is then compared with the immediately preceding and following ISIs. When both flanks are available,

$$S_{\mathrm{pre}}(c) = \frac{\Delta_{a-1}}{Q_{90}(c)}, \qquad a > 2,$$

$$S_{\mathrm{post}}(c) = \frac{\Delta_{b+1}}{Q_{90}(c)}, \qquad b < n,$$

and

$$S_{\min}(c) = \min\{S_{\mathrm{pre}}(c), S_{\mathrm{post}}(c)\}.$$

A two-sided burst-family candidate must exceed the resolved flank-contrast threshold and satisfy bridge and compactness limits. Edge-limited or one-sided candidates can be retained for review as possible_burst rather than discarded. This value is a review status in the current implementation, although the single exported label track requires it to be encoded alongside biological event labels.

Spike count determines the reported burst subtype. The default schema treats 3-10 spikes as classical-burst scale and 11-15 spikes as long-burst scale. Longer compact structures remain prolonged burst-like candidates unless the investigator supplies preparation-specific criteria.

High-frequency spiking, high-frequency tonic and tonic states

High-frequency spiking is modeled as a sustained epoch rather than a burst event. Candidate runs require a large proportion of short ISIs while allowing a limited number of moderate gaps. Evidence includes Q50, Q80, Q90 and Q95, short-ISI fraction, bridge burden, tolerated-gap fraction, large-ISI fraction and the longest sequence of large intervals. The short-ISI fraction is

$$F_{\mathrm{short}}(c) = \frac{1}{N_{\mathrm{ISI}}(c)}\sum_{i=a}^{b}\mathbf{1}\{\Delta_i \le \theta_{\mathrm{short}}\}.$$

Acceptance can follow a strict Q90 route or a robust Q80/majority route. Longer epochs receive span-aware priority, whereas candidates dominated by repeated embedded burst packets can be suppressed so that packetized bursting remains visible.

Tonic and high-frequency tonic candidates require sustained firing within family-specific ISI bands and satisfy regularity constraints based on CV, LV and MM. High-frequency tonic detection also applies a lower ISI floor to avoid reclassifying extreme burst cores as tonic. In version 1.2.1 these categories share a mutually exclusive automatic label track with high-frequency spiking and burst-family events. The resulting cross-family competition is an implementation constraint, not a claim that firing state and firing rate are the same biological dimension.

Pause candidates

Pause candidates are single long-gap intervals. The effective floor combines an absolute threshold, the upper tail of the train-level ISI distribution and guardrails for slow tonic firing. Each long interval is compared with local and global median context so that ordinary slow discharge is less likely to be labeled as a pause.

Weighted interval selection and manual-label policy

Different detector layers can propose overlapping intervals. To produce one public automatic label track, each candidate receives a value V(c) based on family, explicit priority, score and span. Dynamic programming selects the non-overlapping candidate set that maximizes total value:

$$\mathcal{C}_{\text{sel}} = \underset{\mathcal{C}}{\operatorname{argmax}} \sum_{c \in \mathcal{C}} V(c), \qquad c_i \cap c_j = \varnothing \quad \forall i \neq j.$$

The procedure makes cross-family competition deterministic and reproducible. It does not eliminate the conceptual limitation of forcing event type, state and rate labels into a single mutually exclusive track; suppressed and non-selected candidates therefore remain available in the diagnostic output.

Manual labels are stored separately from automatic evidence. With manual protection enabled, a reviewer label remains dominant in final interpretation and export, while automatic candidates can still be generated on eligible intervals for comparison. Explicit negative labels can veto candidate promotion, and the veto is retained in the decision record.

Candidate ledgers and eventness audit

Selected candidates are written to public event and candidate ledgers. Rejected, blocked, profile-only, suppressed and non-selected windows remain in Candidate_diagnostic_audit.csv. Keeping these outputs separate prevents exploratory windows from inflating biological counts while preserving the evidence needed to diagnose failure.

After selection, STPD writes candidate features and final-decision records. The eventness audit summarizes whether a window is more event-like or state-like from clipped edge contrast, context contrast, return-to-baseline evidence and regularity. If Cedge, Cctx and Creturn denote the available components,

$$\text{Eventness}(c) = \text{mean}\left(\min\{C_{\text{edge}}(c), C_{\text{ctx}}(c)\}, C_{\text{return}}(c)\right),$$

with unavailable components omitted. Regularity is calculated from normalized CV, LV and Q90/Q10 evidence. Eventness is a review aid; it is not used as an independent physiological ground truth.

Literature-linked support methods

STPD includes two literature-linked burst-support layers. For the Mean-ISI method (Chen et al., 2009), let Ti denote valid ISIs and $\bar{T}$ their global mean. The threshold is

$$ML = \frac{1}{|I_<|} \sum_{i \in I_<} T_i\,, \qquad I_< = \{i : T_i < \bar{T}\}.$$

Candidate windows are proposed when the mean of consecutive ISIs does not exceed ML. The threshold principle follows the published method; exhaustive window enumeration, computational budgets and merging are implementation choices in STPD.

The LogISI layer follows the threshold-evidence strategy of Pasquale et al. (2010). ISIs are transformed as

$$z_i = \log_{10}(1000 T_i),$$

where Ti is measured in seconds and 1000Ti in milliseconds. A smoothed logISI histogram is searched for peaks and valleys, and void-parameter evidence is used to propose a threshold. If the evidence is insufficient, the threshold remains unresolved rather than being converted into a final label.

Event-level validation against reference intervals

When reference intervals are available, STPD can compare automatic events with synthetic, expert or external annotations. Reference labels used for reported validation should not also determine the reported threshold configuration. The software supports strict label matching and broader family-level matching for review-oriented outputs.

For automatic interval A and reference interval M, STPD calculates temporal intersection, union and intersection-over-union (IoU). Candidate-reference pairs are assigned by descending IoU subject to minimum-overlap and label-compatibility rules. Event precision, recall and F1 are calculated from matched and unmatched events; onset, offset and duration errors are reported separately.

The present manuscript reports interval-level simulator metrics as the primary quantitative benchmark because those labels were available for every scorable ISI. Event-level validation is a package capability, but no independent expert-labeled clinical event set was available for the reported detector-accuracy analysis.

Manuscript preparation

During manuscript preparation, the authors used generative language tools to suggest revisions to wording and organization. The authors independently checked the scientific statements, numerical results, interpretations and final text and accept responsibility for the submitted version.

Results

Traceable software outputs

A standard STPD run proceeds from timestamp import through QC, threshold freezing, candidate generation, feature calculation, interval selection, optional manual review, reference-label comparison and export. The public result is accompanied by the records needed to reconstruct how it was produced.

The exported objects include the pre-detection QC table, effective-threshold table, public candidate ledger, diagnostic candidate audit, candidate-feature table, final-decision audit, eventness audit, parameter report, validation summaries and run metadata (Table 5). Together, these files connect a reported event to the source timestamps, QC status, threshold source, competing candidates and parameter hash.

Table 5 Primary exported artifacts and their interpretive use

Artifact	Interpretive use
QC table	Documents timestamp integrity, duplicate/artifact status and stationarity risks
Threshold table	Records candidate threshold values, selected values and sources
Candidate ledger	Lists public automatic candidates selected for review or reporting
Diagnostic candidate audit	Retains rejected, blocked, suppressed, profile-only and non-selected windows
Final-decision audit	Links candidate evidence, interval selection and manual policies to public labels
Validation summaries	Report agreement with simulator or expert-reference intervals when available
Run metadata	Records software version, input manifest, dependency information and parameter hash

Execution check on a bundled coded clinical timestamp dataset

Package-level execution was checked with the bundled coded clinical CSV file P1_PD-STN_2017.csv using seconds as the input unit and a hard-artifact threshold of 0.0009 s. The file contains de-identified spike timestamps obtained from one participant with Parkinson's disease during standard-of-care deep brain stimulation surgery. P1 is a study code used in place of the participant's identity. Intraoperative microelectrode recording was performed for clinical target localization and electrode-placement guidance and was not added or extended for this software study. The participant's informed consent and authorization from the data-holding medical institution covered the clinical recording, secondary research use, publication of scientific findings and public sharing of the coded timestamp data. The file was used only to test import, QC, detection and export; it was not used for patient-level inference or detector-accuracy claims.

The file contained 23 spike trains and 16,728 timestamps. Recomputed ISIs included 48 intervals shorter than 0.9 ms. These intervals remained in the timestamp and QC outputs but were excluded from valid-ISI evidence. STPD generated the expected event-level and spike/ISI-level exports, including QC flags, public event tables and diagnostic records.

Interval-level benchmark on synthetic spike trains

The benchmark contained 8,776 timestamps and 8,746 scorable ISIs. The class distribution was intentionally imbalanced: high-frequency spiking accounted for 4,381 intervals (50.1%), high-frequency tonic for 1,537 (17.6%), tonic for 1,279 (14.6%), Noisy for 902 (10.3%), Burst for 498 (5.7%) and Pause for 149 (1.7%). Accuracy was therefore treated as a secondary measure and interpreted with macro-F1, class-specific precision and recall, and balanced accuracy.

With automatically resolved thresholds, six-class scoring yielded accuracy = 0.703, inclusive macro-F1 = 0.498 and weighted F1 = 0.669 (Tables 6 and 7). High-frequency spiking was the most stable class (F1 = 0.906). Burst family (F1 = 0.132), tonic (F1 = 0.489), pause (F1 = 0.469) and high-frequency tonic (F1 = 0.355) were less stable. The binary patterned-versus-unclassified weighted F1 was 0.904, but this coarse result did not imply reliable assignment among individual pattern classes.

Simulator-informed thresholds increased accuracy to 0.756 and weighted F1 to 0.712. Inclusive macro-F1 was 0.666 after assigning F1 = 0 to high-frequency tonic, for which the detector emitted no predictions. High-frequency spiking, tonic and pause were recovered with F1 values of 0.922, 0.961 and 1.000, respectively. Burst-family recall increased to 0.940, but precision remained 0.316 (F1 = 0.473). Favorable threshold information therefore changed the operating point without resolving all cross-family competition.

Table 6 Primary interval-level benchmark summaries under the fixed scoring crosswalk

Metric	Automatic	Simulator-informed
Accuracy	0.703	0.756
Macro-F1, all reference classes	0.498	0.666
Macro-F1, emitted classes only	0.498	0.799
Weighted F1, all reference classes	0.669	0.712
Binary patterned vs unclassified weighted F1	0.904	0.907

Table 7 Class-specific interval-level performance on the simulated benchmark

Reference class	n	Auto P	Auto R	Auto F1	Informed P	Informed R	Informed F1
Burst family	498	0.135	0.129	0.132	0.316	0.940	0.473
HF spiking	4,381	0.839	0.984	0.906	0.952	0.893	0.922
HF tonic	1,537	1.000	0.216	0.355	NA	0.000	0.000
Pause	149	0.306	1.000	0.469	1.000	1.000	1.000
Tonic	1,279	0.788	0.355	0.489	0.941	0.981	0.961
Unclassified	902	0.482	0.928	0.635	0.492	0.914	0.640

Values are interval-level precision (P), recall (R) and F1. HF, high-frequency. Simulator-informed thresholds were derived from the simulator configuration and are reported as an oracle-style sensitivity analysis. Precision is undefined when no intervals are predicted; high-frequency tonic F1 was set to 0 for the inclusive macro-average.

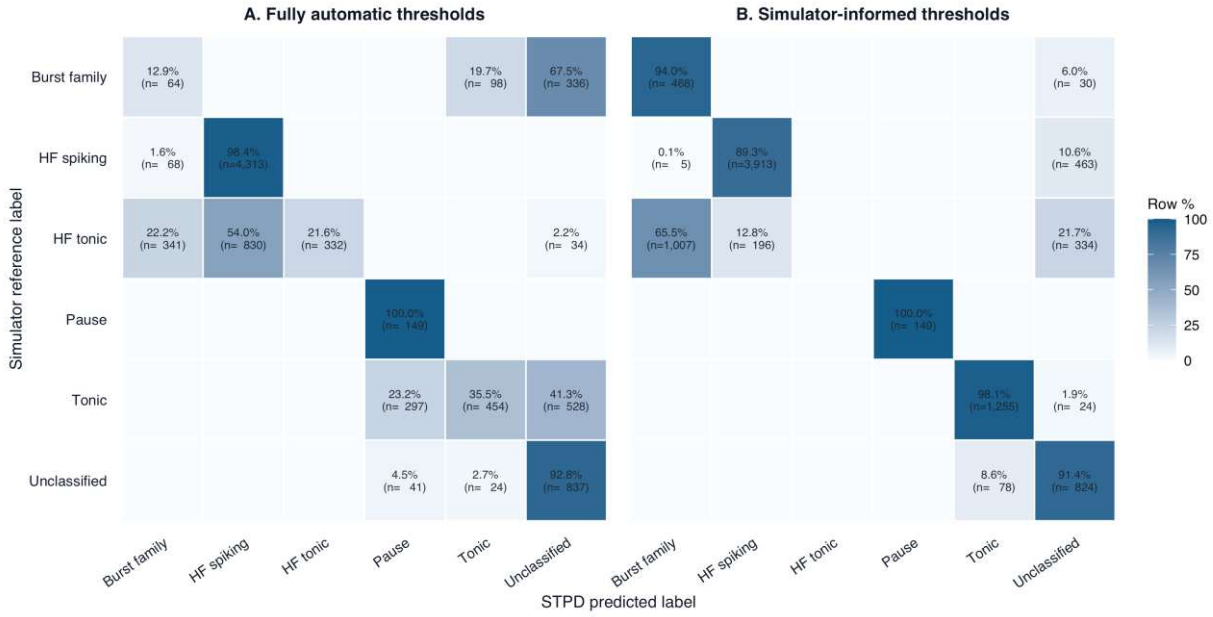


Fig. 4 *Interval-level confusion matrices for the synthetic benchmark. Rows show simulator reference labels and columns show STPD predictions after pooling burst, possible_burst and long_burst as burst family. (A) Automatic thresholds. (B) Simulator-informed thresholds. Noisy reference intervals are compared with the detector's unclassified output as an operational abstention mapping. The absence of high-frequency tonic predictions in panel B shows that threshold information alone did not resolve competition among high-frequency labels.*

Burst-family comparison with Mean-ISI and LogISI

The burst subset illustrates the effect of class imbalance. A classifier that never predicts Burst achieved accuracy = 0.943 because only 498 of 8,746 intervals were Burst. Automatic STPD achieved lower raw accuracy (0.904) but identified 64 Burst intervals; its balanced accuracy was 0.539 and burst-family F1 was 0.132. Accuracy alone was therefore not an informative burst endpoint.

Mean-ISI and LogISI were run in automatic support-layer mode and overcalled Burst intervals, yielding F1 = 0.020 and 0.082, respectively (Table 8). Simulator-informed bands were supplied only to the STPD sensitivity run, not to the support methods. Under that oracle setting, STPD burst recall rose to 0.940 and balanced accuracy to 0.909, while precision remained 0.316. The comparison therefore characterizes operating behavior; it is not a fully calibrated head-to-head benchmark.

Table 8 Burst-family interval detection against simulator Burst labels

Condition	Method	Predicted n	TP/FP/FN	P	R	F1	Bal. acc.	Acc.
Automatic	STPD	473	64/409/434	0.135	0.129	0.132	0.539	0.904
Automatic	Mean-ISI	5,327	57/5270/441	0.011	0.114	0.020	0.238	0.347
Automatic	LogISI	6,314	279/6035/219	0.044	0.560	0.082	0.414	0.285
Automatic	Always-neg.	0	0/0/498	0.000	0.000	0.000	0.500	0.943
Informed	STPD	1,480	468/1012/30	0.316	0.940	0.473	0.909	0.881
Informed	Mean-ISI	5,327	57/5270/441	0.011	0.114	0.020	0.238	0.347
Informed	LogISI	6,314	279/6035/219	0.044	0.560	0.082	0.414	0.285
Informed	Always-neg.	0	0/0/498	0.000	0.000	0.000	0.500	0.943

P, precision; R, recall; Bal. acc., balanced accuracy. The always-negative reference is included to show the inflation of raw accuracy under class imbalance.

Failure modes exposed by the benchmark

The threshold sensitivity analysis separated two sources of error. Improved performance in tonic, pause and burst recall showed that threshold estimation accounted for part of the automatic error. By contrast, the disappearance of high-frequency tonic predictions after simulator-informed calibration showed that favorable thresholds do not guarantee stable final labels when several candidate families compete for the same interval.

Burst-family behavior showed a different trade-off. Automatic thresholds missed most reference Burst intervals, whereas simulator-informed thresholds recovered most of them at the cost of many additional false positives. This pattern is consistent with a screening workflow in which recall and reviewer burden must be chosen explicitly rather than hidden behind a single default score.

The clinical calibration set established only the range of firing rates and ISIs used to construct plausible synthetic inputs. Because the clinical recordings lacked independent event annotations, the present results do not estimate accuracy in the globus pallidus internus or support physiological claims about deep brain stimulation responses.

Discussion

STPD addresses a specific neuroinformatics problem: the final firing-pattern label is often preserved, whereas the decisions that produced it are not. The package makes timestamp QC, threshold provenance, candidate competition and reviewer actions part of the exported result. This contribution is distinct from proposing a universal definition of burst, tonic or pause firing.

The synthetic benchmark shows why this distinction matters. The automatic detector recognized high-frequency spiking well and separated patterned from unclassified intervals at a coarse level, but several fine-grained classes were unstable. Simulator-informed thresholds improved overall performance, yet high-frequency tonic collapsed into competing labels. The main limitation was therefore not simply an absence of plausible thresholds; the structure of the mutually exclusive label track also influenced the result.

A provenance-preserving workflow does not make an inaccurate call correct. It makes the source of the call inspectable. In STPD, a reviewer can determine whether a reference event was never proposed, was rejected by a family rule, lost during interval selection, was blocked by a manual veto or was changed during review. The present manuscript reports the aggregate performance of the current release; stage-specific candidate-recall analyses remain an important next validation step.

Operational interpretation of STPD labels

STPD labels should be read as outputs of explicit operational rules. A burst-family call denotes a compact short-ISI structure with the required local contrast; a pause denotes a long gap relative to absolute and contextual thresholds; tonic and high-frequency calls denote sustained regimes defined by rate and regularity criteria. Their biological meaning remains conditional on recording quality, cell type, brain region, behavioral state and experimental purpose.

The current release places transient events, sustained states, rate classes, review status and abstention on one public label track. This representation is useful for visualization and non-overlapping export, but it can force conceptually different attributes to compete. A future factorized representation should separate event family, state family, rate band, QC status and review status while retaining a deterministic public summary track.

Relationship to classical burst-detection methods

Mean-ISI and LogISI remain useful because they provide literature-linked evidence rather than project-specific labels. Mean-ISI adapts to the train-wide ISI distribution but can be distorted by nonstationarity and long silent periods. LogISI depends on histogram construction, smoothing and peak-valley resolution. STPD keeps these procedures separate from the event grammar so that their assumptions and outputs can be compared instead of combined implicitly.

Relationship to the wider software ecosystem

STPD complements rather than replaces broader neuroinformatics tools. Neo provides an object model for electrophysiology data, Elephant supplies general spike-train and electrophysiology analyses, PySpike focuses on spike-train synchrony and distance measures, and SpikeInterface provides infrastructure for spike sorting and benchmarking (Buccino et al., 2020; Elephant Developers, 2024; Garcia et al., 2014; Mulansky & Kreuz, 2016).

The distinctive layer in STPD is reviewer-facing annotation provenance. A public event call can be linked to input-quality findings, resolved thresholds, candidate evidence, interval-selection decisions, manual-label policies and optional reference overlap. This narrower scope is intended to make firing-pattern annotations easier to inspect and reproduce within preparation-specific workflows.

Limitations

First, the detector uses an explicit rule grammar rather than a canonical biological definition. The rules are inspectable, but their labels remain conditional on the selected parameters and on the current single-track representation.

Second, the primary benchmark was synthetic, finite, imbalanced and scored at the ISI level. Adjacent ISIs within a state are not independent, and interval-level weighting can favor long states. Event-level precision, recall, boundary error and train-level bootstrap intervals should be primary endpoints in subsequent studies.

Third, no independent expert-labeled clinical dataset was available. The clinical recordings were used only for parameter-range calibration, so the reported metrics should not be generalized to a brain region, disease group or recording system without local validation.

Fourth, the Mean-ISI and LogISI comparison used automatic support-layer settings, whereas simulator-derived thresholds were supplied only to the STPD oracle analysis. The comparison reveals overcalling and class-imbalance effects but is not a symmetric calibrated benchmark among competing products.

Finally, the benchmark did not separately quantify candidate generation, rule rejection and interval-arbitration loss. Because STPD records these stages, future validation can report candidate recall, false candidates per minute, selection survival, review burden and final event agreement as distinct outcomes.

Practical use and next validation steps

For current use, investigators should treat STPD as a configurable screening and review environment. The QC table, threshold source, candidate ledger and diagnostic audit should accompany any reported event counts. Thresholds should be fixed before group comparisons, and preparation-specific performance should be estimated against independent annotations or another biological endpoint.

The strongest next evaluation would use independently annotated recordings with train- or subject-level holdout, event-level matching, boundary errors, inter-rater agreement and predefined screening and conservative operating points. Algorithmically, the priority is to separate firing state, rate and review status so that high-frequency tonic activity does not have to compete with tonic or high-frequency spiking as a single flat class.

Conclusion

STPD provides an R/Shiny environment for provenance-aware firing-pattern detection and review. It recomputes ISIs from timestamps, applies QC before detection, freezes and records threshold sources, retains rejected and selected candidates, protects manual labels and exports reproducibility records. The synthetic benchmark identified both a robust high-frequency-spiking regime and clear limitations in burst retrieval and high-frequency-state arbitration.

The software is best used to make preparation-specific annotation decisions visible and testable. Strong biological claims require independent reference labels and event-level validation; STPD supplies the records and comparison tools needed to perform that validation without concealing the assumptions that generated the candidate events.

Statements and Declarations

Ethics approval and consent to participate

All clinical microelectrode-recording data analyzed in this study were acquired during standard-of-care deep brain stimulation procedures for electrophysiological target localization and electrode-placement guidance. No research-specific electrode trajectory, recording session, stimulation procedure or additional invasive intervention was performed. The present work involved secondary analysis of coded or de-identified spike timestamps. Under the applicable institutional policies and determinations of the participating medical institutions, separate ethics committee approval was not required for this secondary analysis. All participants provided informed consent covering clinical recording, secondary research use, publication of scientific findings and sharing of coded or de-identified data. The use and dissemination of the clinical data under the access arrangements described below were also authorized by the respective data-holding medical institutions. Direct personal identifiers were removed before analysis and data sharing.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the China Scholarship Council (grant no. 202307320009).

Author contributions

Houchun Zhou conceived and implemented the software, developed the analysis workflow, performed the reported checks and analyses, and drafted the manuscript. Alexey S. Sedov supervised the work and contributed clinical neurophysiology expertise, interpretation of the microelectrode-recording context and critical revision of the manuscript. Both authors approved the submitted version.

Information Sharing Statement

Software and code availability

SpikeTrainPatternDetector version 1.2.1 is an R/Shiny package with native C components and optional Python integration through reticulate. Source code, installation instructions, parameter files, example data, synthetic-data utilities, validation scripts and reproducibility materials are available from the public repository at https://github.com/zhouhouchun/SpikeTrainPatternDetector-STPD and the archived Zenodo release at https://doi.org/10.5281/zenodo.20650295. Analyses should report the package version or commit, R and operating-system versions, dependency versions, input and parameter hashes, scoring crosswalk, QC table, threshold table and exported audit files.

Data availability

The public software release and Zenodo archive contain the synthetic benchmark materials, analysis scripts, scoring crosswalks, non-clinical result tables and the coded clinical smoke-test file P1_PD-STN_2017.csv. The latter contains spike timestamps obtained from one participant during standard-of-care deep brain stimulation surgery and contains no direct personal identifiers. Public release of this coded file was covered by the participant's informed consent and authorized by the data-holding medical institution. The broader clinical microelectrode-recording dataset and associated task-event annotations are not openly deposited because the data-holding medical institutions retain governance over participant-level clinical datasets. Access requests should be directed to the corresponding author at zhouhouchun@outlook.com. Requests will be reviewed in consultation with the relevant medical institution and may require confirmation of an appropriate research purpose, institutional authorization, a data-use agreement and any necessary ethics authorization. Access is subject to approval by the responsible data custodian.

Acknowledgements

Not applicable.

References

Buccino, A. P., Hurwitz, C. L., Garcia, S., Magland, J., Siegle, J. H., Hurwitz, R., & Hennig, M. H. (2020). SpikeInterface, a unified framework for spike sorting. eLife, 9, e61834. https://doi.org/10.7554/eLife.61834

Chen, L., Deng, Y., Luo, W., Wang, Z., & Zeng, S. (2009). Detection of bursts in neuronal spike trains by the mean inter-spike interval method. *Progress in Natural Science, 19*(2), 229-235. https://doi.org/10.1016/j.pnsc.2008.05.027

Elephant Developers. (2024). *Elephant: Electrophysiology Analysis Toolkit* [Software]. Zenodo. https://doi.org/10.5281/zenodo.1186602

Garcia, S., Guarino, D., Jaillet, F., Jennings, T. R., Proepper, R., Rautenberg, P. L., Rodgers, C. C., Sobolev, A., Wachtler, T., Yger, P., & Davison, A. P. (2014). Neo: an object model for handling electrophysiology data in multiple formats. *Frontiers in Neuroinformatics, 8*, 10. https://doi.org/10.3389/fninf.2014.00010

Holt, G. R., Softky, W. R., Koch, C., & Douglas, R. J. (1996). Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. *Journal of Neurophysiology, 75*(5), 1806-1814. https://doi.org/10.1152/jn.1996.75.5.1806

Mulansky, M., & Kreuz, T. (2016). PySpike: A Python library for analyzing spike train synchrony. *SoftwareX, 5*, 183-189. https://doi.org/10.1016/j.softx.2016.07.006

Pasquale, V., Martinoia, S., & Chiappalone, M. (2010). A self-adapting approach for the detection of bursts and network bursts in neuronal cultures. *Journal of Computational Neuroscience, 29*(1-2), 213-229. https://doi.org/10.1007/s10827-009-0175-1

Shinomoto, S., Miura, K., & Koyama, S. (2005). A measure of local variation of inter-spike intervals. *Biosystems, 79*(1-3), 67-72. https://doi.org/10.1016/j.biosystems.2004.09.023

Zhou, H. (2026). *zhouhouchun/SpikeTrainPatternDetector-STPD: v1.2.1* [Software]. Zenodo. https://doi.org/10.5281/zenodo.20650295

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Stimulationdataresults.zip
- datafortest.zip