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.

# biotrace2csv: Data Extraction Tool for BioTrace+ PDF Data Reports

**Metehan Özsoy**

metehanozsoyacademic@gmail.com

Celal Bayar University https://orcid.org/0000-0001-9145-6537

**Method Article**

**Keywords:** Heart Rate Variability, BioTrace+, Data Extraction, Open Source, Python, Biofeedback, Neurofeedback

**Posted Date:** September 3rd, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10853517/v1

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

**Additional Declarations:** The authors declare no competing interests.

## Abstract

BioTrace+ software generates comprehensive PDF reports for heart rate variability (HRV), physiological channel statistics and HRV resonance frequency assessments, but the manual extraction of these data for further analysis is error-prone and time-consuming. biotrace2csv is a free, open-source command-line tool that automates the parsing of all three report types and exports the data into clean, comma-separated values (CSV) tables. It supports single files, batch processing and automatic type detection. I validated the tool on 20 randomly selected BioTrace+ reports, achieving 100% agreement with manual extraction and reducing processing time from 71.6 seconds per file to 1.45 seconds per file — a speed-up of approximately 50-fold.

## Introduction

Heart rate variability (HRV) biofeedback and physiological monitoring are widely used in psychophysiology, sports science, and clinical research (Shaffer and Ginsberg 2017). BioTrace+ (MindMedia 2018), the software used with the MindMedia Nexus-10 MK II, is a commonly used platform for physiological data acquisition that outputs detailed session reports in PDF format. Despite the richness of these reports, researchers often manually transcribe values into spreadsheet applications, a process that scales poorly and risks introducing transcription errors.

Several commercial and open-source solutions exist for general PDF table extraction, but none are tailored to the specific, non-tabular layout of BioTrace+ reports, which intermix textual metadata, graphical elements, and numerical results. I developed biotrace2csv (Özsoy 2026b) to fill this gap: a lightweight, cross-platform Python tool that faithfully parses the unique structure of BioTrace+ reports and produces analysis-ready CSV files.

## Methods

# Tool Description

biotrace2csv is written in Python 3 and relies on the pdfplumber library for text extraction. It accepts three mutually exclusive mode flags:

- -hrv - Extracts time-domain and frequency-domain HRV metrics (e.g., SDNN, RMSSD, HF, Total Power).
- -pd - Extracts per-segment, per-channel physiological statistics (Min, Max, Mean, Var, StdDev, %>Th1, %<Th2) in long-format CSV.
- -rf - Extracts HRV resonance frequency information (Peak Frequency, Peak Power, Cycles per minute).

Additionally, the -a flag automatically detects the report type of each PDF in a directory and processes them accordingly. All numeric values in the output CSVs use a comma as the decimal separator. Invalid

numeric placeholders (e.g., -1.#J) are replaced by empty cells.

## Example Usage

Processing a folder of mixed report types in quiet mode:

python3 biotrace2csv.py ~/all_reports/ -a -q

## Research Application

I developed and validated the tool during a thesis study investigating autonomic adaptations in novice archers (Özsoy 2026a). Twenty-six university students (20 male, 6 female) participated in a 14-week basic archery training program. I collected physiological recordings (ECG and respiration) during two 4-minute conditions: seated rest with eyes open, and 10-arrow shooting from an 18-meter distance. I administered the experimental protocol twice: after the 5th week (once participants had learned basic shooting and safety) and at the end of the 14th week. Each participant session generated five BioTrace + PDF reports (HRV metrics, channel statistics, and respiratory parameters). In total, I processed 260 PDF files using biotrace2csv, reducing data preparation time from an estimated several hours to a few minutes while ensuring consistency across all participants.

## Results

# Validation against manual extraction

To assess the reliability of biotrace2csv, I manually processed 20 PDF reports from 4 randomly selected participants (5 reports per participant). The following metrics were extracted directly from the PDF files: NNMin, NNMax, NNMean, SDNN, RMSSD, NN50, pNN50, Baevsky's Stress Index (BSI), Total Power (TP), LF, pLF, HF, pHF, LF/HF, Peak Frequency, Cycles, Resonance Power, Heart Rate (HR), Respiration Rate (RSP), and Respiration Amplitude (RSP Amp). The same files were then processed with biotrace2csv for comparison.

Manual extraction of the 20 files required 23 minutes and 52 seconds (approximately 71.6 seconds per file). Extrapolating this rate to the full thesis dataset of 260 files yields an estimated manual processing time of approximately 5 hours and 10 minutes. In contrast, biotrace2csv processed the same 20 files in 28.943 seconds (1.45 seconds per file), corresponding to a full-dataset processing time of approximately 6 minutes and 16 seconds — a speedup of approximately 50-fold. All extracted numerical values matched the corresponding values obtained by manual extraction, including cases where invalid placeholders (e.g., -1.#J) were present and correctly handled by the tool.

The tool generated a total of 304 CSV rows from the 20 validation PDFs, corresponding to an average of 15.2 rows per file. The extracted data were sufficient for all subsequent statistical analyses in the thesis study (Özsoy 2026a).

Example comparison of manual and automated extraction for a single participant's baseline record values.

| Metric | Manual | biotrace2csv |
|---|---|---|
| NNMin (ms) | 629 | 629 |
| NNMax (ms) | 994 | 994 |
| NNMean (ms) | 825.49 | 825.49 |
| SDNN (ms) | 61.2 | 61.2 |
| RMSSD (ms) | 47.8 | 47.8 |
| NN50 (count) | 43 | 43 |
| pNN50 (%) | 15.3 | 15.3 |
| BSI | 72.51 | 72.51 |
| TP (ms²) | 4793 | 4793 |
| LF (ms²) | 1862 | 1862 |
| pLF (%) | 38.84 | 38.84 |
| HF (ms²) | 2389 | 2389 |
| pHF (%) | 49.83 | 49.83 |
| LF/HF | 0.8 | 0.8 |
| Peak Frequency (Hz) | 0.094 | 0.094 |
| Cycles (per min) | 5.6 | 5.6 |
| Resonance Power | 213.3 | 213.3 |
| HR (bpm) | 72.81 | 72.81 |
| RSP (breaths/min) | 18.62 | 18.62 |
| RSP Amp. | 3.33 | 3.33 |

## Discussion

biotrace2csv eliminates manual transcription and the associated human error, significantly speeding up the data pipeline for BioTrace+ users. The validation results confirm that the tool produces identical numerical outputs to manual extraction while dramatically reducing the time required. In the thesis study, I was able to process all 260 PDF reports in a single automated run, eliminating a potential source of human error and reducing the time required for data preparation.

The tool's open-source nature and permanent archival (DOI: https://doi.org/10.17605/OSF.IO/DWFXK) support reproducible research. Current limitations include its exclusive focus on BioTrace + PDF output and the handling of rare artifact notations. Future work will add metadata extraction (session name, date, duration), error logging and channel filtering alongside general code quality and performance improvements.

## Declarations

Ethics Approval Statement: The human participant data used for validating the biotrace2csv tool in this manuscript were collected as part of a previously completed thesis study. That study was reviewed and approved by the Manisa Celal Bayar University Faculty of Medicine Health Sciences Ethics Committee under decision number 20.478.486 dated 15/04/2026. All procedures involving human participants in the original study were conducted in accordance with the ethical standards of the institutional research committee and with the Declaration of Helsinki and its later amendments.

Participant Consent Statement: Informed written consent was obtained from all individual participants included in the original thesis study prior to their participation. Participants were informed about the purpose of the research, the physiological recording procedures, the voluntary nature of their participation, and their right to withdraw at any time without penalty. No new human participants were recruited for the current software tool preprint; only previously collected, fully anonymized data from the approved thesis study were used for validation purposes.

## Availability

- **Project name**: biotrace2csv
- **License**: GNU General Public License v3.0
- **Source code**: https://codeberg.org/metehanozsy/biotrace2csv
- **Archived version (DOI)**: https://doi.org/10.17605/OSF.IO/DWFXK
- **Programming language**: Python 3
- **Operating systems**: Linux, Windows, macOS

## References

MindMedia. 2018. "BioTrace+ Software." https://mindmedia.com/products/biotrace.

Özsoy, Metehan. 2026a. "Autonomic Adaptations in Novice Archers: Task Related Heart Rate Variability During Rest and Shooting." Master's thesis, Manisa: Manisa Celal Bayar University.

———. 2026b. "Biotrace2csv - BioTrace + PDF Data Extractor." https://doi.org/10.17605/OSF.IO/DWFXK.

Shaffer, Fred, and J. P. Ginsberg. 2017. “An Overview of Heart Rate Variability Metrics and Norms.” *Frontiers in Public Health* 5: 258. https://doi.org/10.3389/fpubh.2017.00258.