— LIFECYCLE —
JOURNAL

Multiview Benchmark of Action Unit Detection with the RIKEN Facial Expression Database

Shushi Namba 1, 2

1. Graduate School of Social Sciences, Hiroshima University, Kagamiyama, Higashi-Hiroshima, Hiroshima, 739-8524 Japan
2. Psychological Process Research Team, Guardian Robot Project, RIKEN, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

Published on 11.07.2025
Outcomes Report
DOI: *10.71240/lcyc.386083*

Abstract

For this study, we benchmark four open-source facial action estimators (OpenFace 3.0, OpenFace 2.2.0, Py-Feat, and PyAFAR) on the RIKEN facial expression database (N = 4,800 images), which is annotated by facial intensity (0–5). Using multiple metrics, we evaluated their detection performance. When using the subset of facial actions common across systems, the average Area Under the Curve scores of the three modern systems were comparable (OpenFace 3.0 = .82; Py-Feat = .83; PyAFAR = .82) and higher than OpenFace 2.2.0 (.75). Facial component-wise analyses revealed systematic differences: OpenFace 3.0 performs best for eyebrow raises and mouth opening; OpenFace 2.2.0 better captures frown/tension-related movements; and PyAFAR excels at smile-related facial action units. These findings, which indicate that optimal tool choice is facial-component-dependent, underscore the value of per-action units.

Registration

No registration exists because the study solely compared existing systems using previously collected data.

Materials

Accessible materials are the followings:

https://www.nature.com/articles/s41598-023-49209-8
https://github.com/CMU-MultiComp-Lab/OpenFace-3.0
https://github.com/TadasBaltrusaitis/OpenFace
https://pyafar.org/ https://py-feat.org/pages/intro.html

Data

https://osf.io/r4dwz/overview?view_only=ac97a2f1ebb249b187ca25de9e8702a1

Code

We cannot deposit the per-system execution pipelines because they depend on user-specific computing environments. However, we have released the analysis code that computes all benchmark metrics from precomputed system outputs.
https://osf.io/r4dwz/overview?view_only=ac97a2f1ebb249b187ca25de9e8702a1

Paper

Multiview Benchmark of Action Unit Detection with the RIKEN Facial Expression Database

By: Shushi Namba

1. Introduction

Facial movements play a central role in coordinating social interaction, communicating not only affect but also thoughts, appraisals, intentions, and action tendencies (Fernández-Dols & Russell, 2017; Mandal & Awasthi, 2015). To provide a theory-neutral, anatomically grounded description of the visible movements which constitute facial behavior, Ekman and colleagues developed the Facial Action Coding System (FACS: Ekman, et al., 2002): this observer-based taxonomy minimizes inference by cataloging facial movements without assigning psychological meaning. Complex expressions are decomposed into combinations of elementary facial actions: Action Units (AUs). Each AU is linked to the contraction of specific muscles. For example, activation of the zygomaticus major, which elevates the lip corners, is coded as AU12, whereas contraction of the orbicularis oculi (pars orbitalis), which raises the cheeks, is coded as AU6. The AUs analyzed for this study are presented in Table 1. Manual FACS annotation has underpinned a large body of research experimentation conducted to elucidate facial behavior and its psychological correlates (Ekman & Rosenberg, 2005).

1. TABLE I Targeted AUs

Action Unit	**FACS Name (Description)**
AU1	Inner brow raiser
AU2	Outer brow raiser
AU4	Brow lowerer
AU5	Upper lid raiser
AU6	Cheek raiser
AU7	Lid tightener
AU9	Nose wrinkler
AU10	Upper lip raiser
AU12	Lip corner puller
AU14	Dimpler

AU17	Chin raiser
AU23	Lip tightener
AU24	Lip pressor
AU25	Lips part
AU26	Jaw drop

Manual FACS annotation is labor-intensive, motivating extensive work on automatic AU detection (Ertugrul et al., 2020). Several open-source toolkits are used widely for this purpose, including OpenFace 2.2.0 (Baltrušaitis et al., 2015; Baltrušaitis et al., 2018), PyFeat (Cheong et al., 2023), and the Automated Facial Affect Recognition toolbox (AFAR: Ertugrul et al., 2019). In evaluations conducted in 2021, OpenFace 2.2.0 outperformed AFAR and PyFeat on our benchmarks (Namba et al., 2021; Namba et al., 2021). The ecosystem has evolved rapidly since then. Also, AFAR has been redesigned with updated landmark estimation and inference pipelines as PyAFAR (Hinduja et al., 2023). Moreover, PyFeat has undergone active development, with notable changes relative to its 2021 version (e.g., the default AU classifier was switched from SVM to XGBoost in December 2022). Since its release, OpenFace 3.0 has reportedly surpassed OpenFace 2.2.0 in terms of performance (Hu et al., 2025).

Given these developments, identifying which system currently offers the most effective AU detection from an end-user's perspective is crucially important. Accordingly, this study assesses these toolkits systematically across multiple criteria using a database with expert manual AU annotations: the RIKEN Facial Expression Database (Namba et al., 2023).

2. Methods

Details of materials, procedures, measures, and analysis are described with sufficient clarity to support reproduction of our work.

2.1 The RIKEN Facial Expression database

The RIKEN Facial Expression Database comprises facial displays from 48 Japanese participants. Each participant produced expressions for 25 personally experienced events spanning affective valence and arousal. For a subset of five event types performed by eight expressers (four women and four men), manual AU annotations were provided by a FACS-certified coder. Annotations cover 29 AUs, including lateralized codes (e.g., AU1Left, AU1Right). Available AUs were the following: 1LR, 2LR, 4LR, 5LR, 6LR, 7LR, 9, 10, 12LR, 14LR, 15LR, 17, 20LR, 23, 24, 25, 26, 27, and 43. Because the evaluated toolkits do not output lateralized AUs, we collapsed laterality by taking, for each AU, the

maximum of the left and right intensities; this choice prioritizes sensitivity to unilateral activations. Each clip includes 120 frames. Therefore, 4,800 frames were analyzed in all.

2.2 OpenFace 2.2.0 and 3.0

We used OpenFace 2.2.0's default CE-CLM (Zadeh et al., 2017) landmarking pipeline. The AU occurrence and intensity are estimated from dimensionality-reduced HOG appearance features in combination with geometric shape features via linear-kernel support vector machines. The pre-trained AU classifiers were trained on widely used corpora (e.g., SEMAINE, BP4D, DISFA, FERA2011, UNBC-McMaster). For our setup, the available AUs were 17: 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, and 45.

OpenFace 3.0 is a multitask deep architecture that integrates modern face detection and alignment modules (Deng et al., 2020; Zhou et al., 2023) and which performs joint inference over a unified facial representation. Its AU head refines predictions using feature-level relations among AUs (e.g., graph-based propagation), which improves its accuracy and efficiency relative to OpenFace 2.2.0 (Hu et al., 2025). For our setup, the available AUs were eight: 1, 2, 4, 6, 9, 12, 25, and 26.

2.3 PyAFAR

PyAFAR is an open-source Python toolbox for automated AU analysis. Its pipeline includes: (i) face detection and individualized tracking (MediaPipe for dense landmarks [Lugaresi et al. 2019], Facenet for re-identification [Schroff et al., 2015], with head pose estimated via a PnP solution [Rocca et al., 2014]); (ii) face normalization using dlib; and (iii) AU occurrence and intensity estimation using ResNet–50 models pre-trained on ImageNet and fine-tuned on BP4D+. The released occurrence models cover 12 AUs. For this study, we analyzed AU occurrences for the following: 1, 2, 4, 6, 7, 10, 12, 14, 15, 17, 23, and 24.

2.4 Py-Feat

Py-Feat is an open-source Python toolkit that spans detection, preprocessing, analysis, and visualization of facial expression data via a Detector and a "Fex" data module, enabling streamlined workflows from frames to statistical summaries and plots. In the default configuration for low-level detectors, Py-Feat includes widely used face and landmark models (e.g., RetinaFace for face detection [Deng et al., 2020]; MobileFaceNet for 68-point landmarks [Chen et al., 2018]), which provide inputs to downstream AU estimation. With emphasis on AU detection, Py-Feat implements

gradient-boosted trees (“Feat-XGB”) classifier to predict 20 AUs: AU 1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 14, 15, 17, 20, 23, 24, 25, 26, 28, and 43.

2.5 Evaluation Metrics

We used two values: AU occurrence (probability of AUs; OpenFace3.0, PyAFAR, Py-feat) and AU intensity (0–5; OpenFace2.2.0). Ground-truth labels were derived from manual FACS annotations that included lateralized AUs. Because no evaluated systems output left/right laterality, we collapsed laterality by taking, for each AU and frame, the maximum of left and right intensities. Binary occurrence labels were then defined as intensity greater than 0. To contextualize class imbalance, this report describes the positives (base rate) per AU as the proportion of frames labeled present in the evaluation set. To reduce instability under severe class imbalance, our primary analyses are restricted to AUs with base rates greater than or equal to.05 (i.e., at least 5% positive frames); rarer AUs are omitted from the main results.

For each system, we consumed its native continuous outputs whenever available to avoid thresholding artifacts. Outputs were evaluated using Pearson’s *r,* Spearman’s ρ, Concordance Correlation Coefficient (CCC), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Quadratic-Weighted Kappa (QWK, 0–5). For QWK, continuous predictions were discretized by rounding to the nearest integer on the 0–5 scale. Pearson’s *r,* which quantifies the strength of a linear association between predictions and ground truth, is invariant to location/scale shifts. Spearman’s ρ captures monotonic association based on ranks, reducing sensitivity to outliers and nonlinearity. The Concordance Correlation Coefficient (CCC) evaluates absolute agreement by combining correlation with penalties for bias and scale differences relative to the identity line. Mean absolute error (MAE) and root mean squared error (RMSE) summarize error magnitudes in the native 0–5 units: MAE averages absolute deviations, whereas RMSE (the square root of mean squared error) emphasizes larger errors. Quadratic-Weighted Kappa (QWK, 0–5), which is computed after rounding predictions to the nearest integer, is a chance-corrected measure of agreement for ordered categories that assigns larger penalties to greater category discrepancies. For *r,* ρ, CCC, and QWK, higher values indicate better performance (maximum = 1), whereas, for MAE and RMSE, lower values indicate better performance (minimum = 0).

3. Results

Table 2 shows average values on the subset of AUs common across systems: AU 1, 2, 4, 6, and 12. Bold entries indicate the best-performing value among systems for each AU–metric comparison. The average AUC scores of the three modern systems were comparable (OpenFace 3.0 =.82, Py-Feat =.83, PyAFAR =.82) and higher than OpenFace 2.2.0 (.75). With the exception of AUC, OpenFace 3.0 achieved the best overall performance among the four systems, yielding the highest association and agreement scores (Pearson’s *r,* Spearman’s ρ, CCC, QWK) and the lowest

deviation/error magnitudes (MAE, RMSE). PyAFAR and Py-Feat followed, whereas OpenFace 2.2.0 performed worse on the intersection of AUs shared across systems. Per-AU and per-system results for all evaluation metrics are available via our OSF repository (link: *https://osf.io/r4dwz/?view_only = ac97a2f1ebb249b187ca25de9e8702a1*). One contributing factor is its markedly low accuracy for eyebrow-raising AUs (AU1/2): OpenFace 2.2.0 achieved AUCs of.60 (AU1) and.62 (AU2). Aside from PyAFAR's AU4 (AUC =.60), these were the only AUs with AUC <.70. Taken together, OpenFace 2.2.0 appears to struggle with upward eyebrow movements, whereas PyAFAR apparently has difficulty with brow-lowering/frowning (AU4).

1. TABLE II Average Metrics Among Common AUs

System	Pearson	Spearman	CCC	MAE	RMSE	QWK	AUC
OpenFace 2.0	.50	.44	.29	.32	.62	.42	.75
OpenFace 3.0	*.67*	*.51*	*.61*	*.12*	*.23*	*.61*	.82
PyAFAR	.51	.44	.45	.12	.24	.44	.82
Py-Feat	.59	.47	.43	.24	.28	.43	*.83*

Although OpenFace 3.0 achieved the highest overall performance on the set of AUs common to all toolkits, system-specific strengths emerged at the level of individual AUs. OpenFace 2.2.0 was strong for AU4 (*r* =.60, AUC =.77), AU7 (*r* =.68, AUC =.85), and AU17 (*r* =.58, AUC =.86). By contrast, PyAFAR excelled for AU6 (*r* =.80, AUC =.83) and AU12 (*r* =.77, AUC =.87). Py-Feat did not exhibit a clear AU for which it outperformed the others. Consistent with its overall strength, OpenFace 3.0 showed particularly strong results for AU1 (*r* =.65, AUC =.82) and AU2 (*r* =.75, AUC =.88). It was markedly superior for AU25 (*r* =.85, AUC =.96).

With respect to AUs often discussed in relation to emotion categories, AU5 (upper-lid raiser; "surprise") is produced only by OpenFace 2.2.0 (*r* =.13; AUC =.59) and Py-Feat (*r* =.25; AUC =.66) in our setup. Its performance was modest for both. AU9 (nose wrinkler; "disgust") is available from OpenFace 2.2.0, OpenFace 3.0, and Py-Feat; here OpenFace 3.0 performed best, but its absolute accuracy was lower than for the other described AUs (*r* =.24, AUC =.71).

4. Discussion

This study provides a contemporary evaluation of four open-source AU detection systems on an expert-annotated dataset. On the subset of AUs shared across systems, the three modern toolkits (OpenFace 3.0, PyAFAR and Py-Feat) exhibited comparable discrimination (AUC ≈.82—.83), clearly exceeding OpenFace 2.2.0 (AUC =.75). Beyond discrimination, however, OpenFace 3.0 consistently achieved stronger association and agreement with ground truth (higher *r,* ρ, CCC, QWK) and lower error magnitudes (MAE, RMSE), indicating not only that it separates positive from negative frames

well, but also that its continuous scores track annotated intensity more faithfully. This pattern underscores the value of using a multi-metric lens: AUC can mask differences in calibration and intensity sensitivity that are captured by correlation, concordance, and error metrics.

At the AU level, system-specific strengths emerged despite the overall strength of OpenFace 3.0. It was particularly effective for eyebrow raises (AUs 1, 2) and for mouth opening (AUs 25, 26), whereas OpenFace 2.2.0 performed well on brow tension/frown-related units (AUs 4, 7, and 17), and PyAFAR excelled on smile-related units (AUs 6, 12). These differences probably reflect design choices: OpenFace 3.0's modern detection/alignment (including STAR [Zhou et al., 2023]) and joint inference appear to capture small vertical displacements (eyebrow raises) and aperture changes (mouth opening) better. OpenFace 2.2.0's HOG+linear pipeline retains competitive sensitivity for configurations dominated by shape/tension cues (e.g., AU4, 7, 17). Also, PyAFAR's strengths align with smile-relevant musculature, which is consistent with its fine-tuning data which included smile-elicitation-conditions (Namba et al., 2022). Together, these findings contraindicate a single "best" toolkit and favor AU-dependent tool selection.

Methodologically, the study highlights that no single metric suffices. Discrimination (AUC) is necessary but not sufficient. Association (Pearson's *r,* Spearman's ρ) and agreement (CCC, QWK) reveal whether systems track the graded nature of AU intensity. Error metrics (MAE, RMSE) capture practical deviations in the native 0–5 scale. For practitioners, the current study recommended choosing of a toolkit with demonstrated strength on the target AUs rather than defaulting to a single package: for example, OpenFace 3.0 for AU1/2/25/26, PyAFAR for AU6/12, and OpenFace 2.2.0, for which AU4/7/17 are crucially important.

Several limitations must be considered to temper generalization of this study's findings. First, the RIKEN database used for this study includes posed expressions from Japanese participants. Also, crucially, our benchmark restricts analysis to near-frontal views. As a result, we did not assess robustness to head-pose variation (yaw/pitch/roll) or camera angle. Performance in off-axis conditions might therefore diverge from what this study shows. Second, AU annotations were available for a subset of event types and expressers. Then we collapsed lateralized AUs by taking the maximum of left–right intensities. This choice prioritizes sensitivity to unilateral activations, but it also compresses subtle asymmetries and precludes evaluation of side-specific performance. Third, the toolkits expose heterogeneous outputs (some provide occurrence probabilities, others provide continuous intensity). For that reason, despite evaluation of all systems on continuous outputs against a common 0–5 reference, residual scale mismatches might influence agreement and error metrics. Finally, we did not probe robustness to occlusion or illumination changes, nor did we evaluate cross-database transfer.

Future work should therefore extend evaluation to systematic head-pose manipulations (e.g., controlled bins of yaw/pitch/roll) and in-the-wild footage to quantify angle robustness and ecological validity. A priority for model development is to produce lateralized AU outputs. This production will require training on datasets that include left/right AU annotations, including the

lateralized labels available in RIKEN or FEAFA+ (Gan et al., 2022). Moreover, it will require designing inference heads that model side specificity explicitly. Beyond lateralization and pose, incorporating temporal models that leverage dynamics rather than frame-wise inference, exploring per-AU ensembling or calibration to combine complementary strengths across toolkits, and assessing fairness/subgroup performance across age, gender, and ethnicity are expected to strengthen practical deployment and scientific conclusions further.

Supporting Information

Supporting information files can be found here: *https://osf.io/25z6d/files/w3jvs*

References

Baltrušaitis, T., Mahmoud, M., & Robinson, P. (2015). Cross-dataset learning and person-specific normalisation for automatic action unit detection. In Proc. 11th IEEE Int. Conf. and Workshops on Automatic Face and Gesture Recognition (FG) (pp. 1–6). Ljubljana, Slovenia. DOI: 10.1109/FG.2015.7284869.

Baltrušaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). OpenFace 2.0: Facial behavior analysis toolkit. In Proc. 13th IEEE Int. Conf. Automatic Face & Gesture Recognition (FG) (pp. 59–66). Xi'an, China. DOI: 10.1109/FG.2018.00019.

Chen, S., Liu, Y., Gao, X., & Han, Z. (2018). MobileFaceNets: Efficient CNNs for accurate real-time face verification on mobile devices. arXiv:1804.07573. DOI: 10.48550/arXiv.1804.07573.

Cheong, J. H., Jolly, E., Xie, T., Byrne, S., Kenney, M., & Chang, L. J. (2023). Py-feat: Python facial expression analysis toolbox. Affective Science, 4(4), 781–796. DOI: 10.1007/s42761–023–00145–2.

Deng, J., Guo, J., Ververas, E., Kotsia, I., & Zafeiriou, S. (2020). RetinaFace: Single-shot multi-level face localisation in the wild. In Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR) (pp. 5203–5212). DOI: 10.1109/CVPR42600.2020.00525.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). Facial Action Coding System (2nd ed.). Salt Lake City, UT: Research Nexus eBook.

Ekman, P., & Rosenberg, E. L. (Eds.). (2005). What the face reveals (2nd ed.). New York, NY: Oxford University Press. DOI: 10.1093/acprof:oso/9780195179644.001.0001

Ertugrul, I. O., Cohn, J. F., Jeni, L. A., Zhang, Z., Yin, L., & Ji, Q. (2020). Crossing domains for AU coding: Perspectives, approaches, and measures. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2, 158–171. DOI: 10.1109/TBIOM.2020.2977225.

Ertugrul, I. O., Jeni, L. A., Ding, W., & Cohn, J. F. (2019, May). AFAR: A deep learning based tool for automated facial affect recognition. In 2019 14th IEEE Int. Conf. Automatic Face & Gesture Recognition (FG 2019) (p. 1). DOI: 10.1109/FG.2019.8756623.

Fernández-Dols, J.-M., & Russell, J. A. (Eds.). (2017). The science of facial expression. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780190613501.001.0001.

Gan, W., Xue, J., Lu, K., Yan, Y., Gao, P., & Lyu, J. (2022). FEAFA+: An extended well-annotated dataset for facial expression analysis and 3D facial animation. In Proc. SPIE 12342, 14th Int. Conf. Digital Image Processing (ICDIP 2022) (pp. 307–316). DOI: 10.1117/12.2643588.

Hinduja, S., Ertugrul, I. O., Bilalpur, M., Messinger, D. S., & Cohn, J. F. (2023). PyAFAR: Python-based automated facial action recognition library for use in infants and adults. In 2023 11th Int. Conf. Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW) (pp. 1–3). DOI: 10.1109/ACIIW58564.2023.10389135.

Hu, J., Mathur, L., Liang, P. P., & Morency, L.-P. (2025). OpenFace 3.0: A lightweight multitask system for comprehensive facial behavior analysis. arXiv:2506.02891.

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., … & Grundmann, M. (2019). MediaPipe: A framework for perceiving and processing reality. Proc. Third Workshop on Computer Vision for AR/VR at IEEE CVPR. DOI: 10.48550/arXiv.1906.08172.

Mandal, M. K., & Awasthi, A. (Eds.). (2015). Understanding facial expressions in communication: Cross-cultural and multidisciplinary perspectives. Berlin/Heidelberg: Springer. DOI: 10.1007/978–81–322–1934–7.

Namba, S., Sato, W., & Matsui, H. (2022). Spatio-temporal properties of amused, embarrassed, and pained smiles. Journal of Nonverbal Behavior, 46(4), 467–483. DOI: 10.1007/s10919–022–00404–7.

Namba, S., Sato, W., Namba, S., Nomiya, H., Shimokawa, K., & Osumi, M. (2023). Development of the RIKEN database for dynamic facial expressions with multiple angles. Scientific Reports, 13, 21785. DOI: 10.1038/s41598–023–49209–8.

Namba, S., Sato, W., Osumi, M., & Shimokawa, K. (2021). Assessing automated facial action unit detection systems for analyzing cross-domain facial expression databases. Sensors, 21(12), 4222. DOI: 10.3390/s21124222.

Namba, S., Sato, W., & Yoshikawa, S. (2021). Viewpoint robustness of automated facial action unit detection systems. Applied Sciences, 11(23), 11171. DOI: 10.3390/app112311171.

Rocca, F., Mancas, M., & Gosselin, B. (2014). Head pose estimation by perspective-n-point solution based on 2D markerless face tracking. In Intelligent Technologies for Interactive Entertainment

(INTETAIN 2014) (pp. 67–76). Cham: Springer. DOI: 10.1007/978–3–319–08189–2_8.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. In 2015 IEEE Conf. Computer Vision and Pattern Recognition (CVPR) (pp. 815–823). DOI: 10.1109/CVPR.2015.7298682.

Zadeh, A., Chong, L. Y., Baltrušaitis, T., & Morency, L.-P. (2017, Oct. 22–29). Convolutional experts constrained local model for 3D facial landmark detection. In Proc. IEEE Int. Conf. Computer Vision Workshops (ICCVW) (pp. 2519–2528). Venice, Italy. DOI: 10.1109/ICCVW.2017.296.

Zhou, Z., Li, H., Liu, H., Wang, N., Yu, G., & Ji, R. (2023). Star loss: Reducing semantic ambiguity in facial landmark detection. In Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR) (pp. 15475–15484). DOI: 10.1109/CVPR52729.2023.01485.

Declarations

Ethics

This work is a technical validation study that analyzes outputs of existing software on previously collected, consented datasets: no new data were collected; no human subjects were recruited; and no evaluation of individual traits or attributes was performed. The data used were obtained under their original licenses/consents and were handled in de-identified form with only aggregate results and non-identifying IDs reported. IRB/ethics approval: N/A

Competing Interests

The authors declare that no conflicts of interest exist.

Funding

This work was supported by JSPS KAKENHI Grant Number JP23K17180. The funder played no role in the study design, data collection, analysis, decision to publish, or submission preparation.

Author Contributions

All CRediT roles were performed solely by the author: Conceptualization; Methodology; Software; Validation; Formal analysis; Investigation; Resources; Data curation; Writing – Original Draft; Writing – Review & Editing; Visualization; Project administration; Supervision; Funding acquisition