Blending Ensemble Learning for Enhanced Arrhythmia Classification Utilizing 12-Lead ECGs

Hai-Chau Le
Department of Data Engineering
Posts and Telecommunications Institute of Technology
Hanoi, Vietnam
chaulh@ptit.edu.vn

Hai-Long Nguyen
Data and Intelligent Systems Laboratory
Posts and Telecommunications Institute of Technology
Hanoi, Vietnam
longnh.b23kd038@stu.ptit.edu.vn

Huu Cam Nguyen
Department of Data Engineering
Posts and Telecommunications Institute of Technology
Hanoi, Vietnam
camnh@ptit.edu.vn

Chien Trinh Nguyen
Department of Signals and Systems
Posts and Telecommunications Institute of Technology
Hanoi, Vietnam
trinhnc@ptit.edu.vn

***Abstract*—The rise in heart-related diseases has driven the development of advanced techniques for identifying irregular heart conditions, and with the progress in artificial intelligence and signal processing, automated arrhythmia classification using machine learning and electrocardiograms (ECG) has become increasingly effective and widely utilized by healthcare professionals. This paper presents an efficient machine learning solution for robust arrhythmia classification using 12-lead electrocardiograms (ECGs) by leveraging blending ensemble learning. Our developed blending model utilizing six base models—Adaptive Boosting (ADA), Decision Trees (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGB)—with a Logistic Regression (LR) meta-model, demonstrates enhanced efficiency in classifying arrhythmias based on 12-lead ECGs. This approach not only exploits the unique strengths of each base model but also captures diverse predictive patterns, which are combined by the LR meta-model into a refined and cohesive output. The use of LR as the meta-model enhances interpretability and generalization, reducing the risk of overfitting and optimizing overall performance. Experimental results demonstrate that the proposed blending ensemble outperforms conventional notable works in terms of accuracy and offers a robust and effective solution for accurate arrhythmia classification, i.e. up to 97.2% accuracy, supporting clinical decision-making.**

***Index Terms*—Machine learning, Ensemble learning, Blending, Electrocardiogram, Arrhythmia classification**

I. Introduction

Cardiovascular diseases are among the most dangerous health issues, responsible for a significant number of deaths globally. According to the WHO, cardiovascular diseases caused 17.9 million deaths in 2019, accounting for 32% of all global deaths, with 85% resulting from heart attacks and strokes [1]. While cardiovascular-related death rates have declined in developed countries due to improved healthcare and healthier lifestyles, they remain high in low- and middle-income countries [2]. Thus, developing effective solutions for early identification of irregular heart rhythms is essential to reduce the impact of these diseases, and the electrocardiogram (ECG) is widely regarded as the most reliable tool for detecting arrhythmias [3]–[5].

The ECG is widely used for cardiac arrhythmia monitoring due to its ability to trace the heart's electrical activity, and recent research has focused on enhancing arrhythmia detection using machine learning (ML) and deep learning techniques [6]–[8]. Studies have utilized various algorithms, such as SVM, KNN, and RF, often in combination or as hybrid models, to improve detection rates of irregular heartbeats [2], [3], [9]. These efforts highlight the ongoing trend of leveraging advanced machine learning techniques to improve the accuracy and efficiency of arrhythmia classification from ECG data.

Moreover, as the clinical standard, thanks to the advantages of offering a more comprehensive view of the heart's electrical activity by capturing signals from 12 different angles and enhancing the ability to diagnose complex arrhythmias, myocardial infarctions, and other cardiac conditions with greater accuracy and specificity, 12-lead ECGs are crucial for thorough cardiac assessments, allowing for more precise and reliable diagnoses than other simplified ECG configurations [5], [10]. Recently, there have been many research attempts to classify arrhythmia based on 12-lead ECGs [11], [12]. Authors in [13] proposed a convolutional neural network (CNN) method to transform ECG signals into RGB images using Continuous Wavelet Transform (CWT) and CNN is used next to extract features out of images to generate a CNN feature representation in which the last layer is soft-max regression for multiclass classification. Especially, the work of [14] developed advanced 12-lead ECG signal processing techniques and two arrhythmia classification methods that are support vector machines and MLP. It's shown that MLP performs better than SVM for the variety of the 12-lead ECG signals with an accuracy of 84.2%. Another impressive work introduced in [15] demonstrated a high accuracy of 97.0% can be achieved for the classification of 4 arrhythmia classes. However, besides the ECG features, they both need further clinical features which may cause difficulties in applying to automated classification systems.

On the other hand, among recently advanced machine learning approaches, blending, an advanced ensemble learning technique, enhances predictive performance by combining the strengths of diverse models through a meta-model, which makes final predictions based on the outputs of multiple base models [16], [17]. The meta-model is trained on a separate validation set to prevent overfitting and ensure generalizability. Popularized in competitions like the Netflix Prize and Kaggle, blending is valued for its versatility and effectiveness in complex predictive tasks. Although blending has been widely applied across various fields, there is currently no known application of this technique for 12-lead ECG arrhythmia classification.

In this paper, we propose an efficient machine learning solution for arrhythmia classification exploiting blending ensemble learning. The developed model combines six base algorithms—ADA, DT, KNN, LR, RF, SVM, and XGB—with a Logistic Regression (LR) meta-model to enhance classification accuracy and generalization. By leveraging the strengths of each base model and refining predictions through the LR meta-model, the solution reduces overfitting and optimizes performance. Experimental results show that this blending ensemble achieves up to 97.2% accuracy, outperforming conventional methods and providing a robust tool for clinical decision-making in arrhythmia detection.

II. Data Processing

In our study, the Chapman ECG dataset—a comprehensive resource developed with Chapman University, Shaoxing People's Hospital, and Ningbo First Hospital—is used as the primary dataset for arrhythmia classification [15]. It includes 10-second, 12-lead ECG recordings from 10,646 patients, sampled at 500 Hz, featuring 11 common cardiac rhythms and 67 additional cardiovascular conditions, all annotated by expert clinicians. The dataset also provides denoised signals and serves as a valuable resource for developing, comparing, and optimizing statistical and machine learning methods in arrhythmia and cardiovascular disorder research.

With the target of an automatic cardiac arrhythmia classification for 12-lead ECG signals, all clinical and ECG-non-related features, i.e. age and gender, are ignored. Signal noise reduction and feature extraction of each lead in the ECG data were performed by utilizing Neurokit2, a Python Toolbox for neurophysiological signal processing [18], and 20 features, including R peaks count, mean, median, stander, range, skewness, and kurtosis of RR interval, mean of R peaks amplitude compared to the isoelectric line, P peaks count over R peaks count, a median of R intervals subtract P intervals, T peaks count, Q peaks count, mean, a variance of QT intervals, PR intervals, PR segment, and ST segment, are extracted from each ECG lead from the original dataset. The worthless features that have too much empty value are dropped.

Moreover, based on clinical relevance, 11 distinct rhythms, as labeled by professional experts, were consolidated into four groups: AFIB (Atrial Fibrillation and Atrial Flutter), SB (Sinus Bradycardia), SR (Sinus Rhythm) and GSVT (General Supraventricular Tachycardia) [11]. Specifically, AFIB encompassed atrial fibrillation and atrial flutter; SB included only sinus bradycardia; SR included both sinus rhythm and sinus irregularity; while GSVT comprised supraventricular tachycardia, atrial tachycardia, atrioventricular node reentrant tachycardia, atrioventricular reentrant tachycardia, and wandering atrial pacemaker. These four grouped labels were employed for the training and evaluation of our models.

Lastly, the ECG feature values are normalized using the MinMaxScaler algorithm, scaling all features to the range [0, 1]. The resulting dataset, comprising 7,449 samples with a total of 240 features, is partitioned into Training and Testing subsets, with proportions of 80% and 20%, respectively. It is important to note that the class distribution ratios of the four arrhythmia classes are preserved across both subsets.

III. Methodology

Figure 1 describes the workflow of our proposed blending ensemble learning approach for 12-lead ECG-based multiclass arrhythmia classification. Blending models, also known as ensemble blending, is a technique in machine learning where multiple models, often of different types, are combined to improve predictive performance compared to individual models [16]. This approach builds on the principles of ensemble learning, which aims to reduce variance, bias, or improve predictions by combining the strengths of various models. While blending is closely related to other ensemble techniques like stacking, bagging, and boosting, it is uniquely characterized by its approach to combining model outputs through a meta-model. Our developed blending model, which combines the predictive capabilities of multiple base models to enhance overall classification performance, integrates six base models and one meta-model, selected from seven fundamental machine learning algorithms: Adaptive Boosting (ADA), Decision Trees (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGB). Each base model is trained independently on the normalized ECG dataset, and their predictions are used to create a new dataset for a meta-model. The meta-model, also selected from the same set of algorithms, is trained to optimally combine the base models' outputs, aiming to improve the classification accuracy and robustness. In addition, the grid search procedure, incorporating 3-fold cross-validation, is employed to select the optimal configuration. In the final step, the meta-model makes the overall arrhythmia predictions based on the base models' outputs from the Testing dataset. This ensemble method is evaluated using standard metrics, ensuring that the class distribution is preserved and the model effectively handles the complexity of multiclass arrhythmia classification, providing a more accurate and reliable diagnostic tool.

A. *Fundamental Machine Learning Algorithms*

To leverage the advantages of the blending ensemble learning approach, seven basic machine learning algorithms that are popular, diverse, and efficient are considered. The selected algorithms are briefly explained as follows.

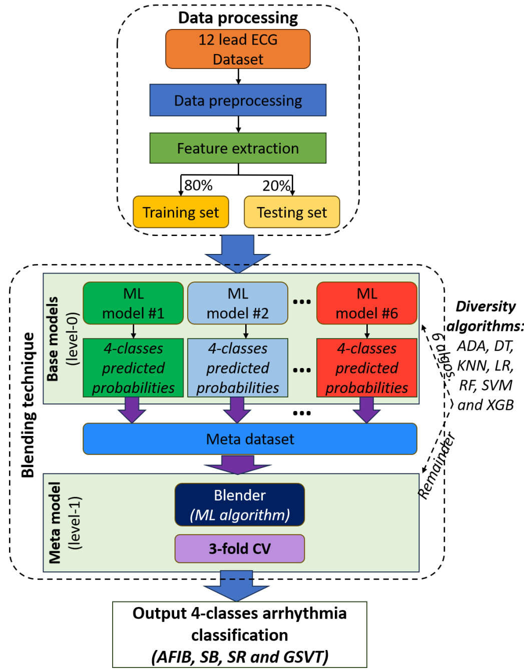

Fig. 1. Workflow of the proposed blending multi-class arrhythmia classification for 12-lead ECGs

- AdaBoost (ADA) is an iterative ensemble learning algorithm for classification that trains base learners on weighted data, increasing the weights of misclassified points in each iteration, and makes final predictions based on a weighted combination of all learners, known for its high accuracy and versatility with various base learners.
- Decision Tree (DT) is a supervised learning algorithm for classification and regression that splits data into branches based on features to form a tree-like model, offering interpretability but prone to overfitting on complex or noisy data.
- k-Nearest Neighbors (KNN) is a non-parametric algorithm that predicts outcomes based on the labels or values of the k closest data points to a query point, using proximity in the feature space.
- Logistic Regression (LR) is a supervised learning algorithm that uses the logistic (sigmoid) function to map inputs to probabilities for classification, relying on assumptions about the data distribution.
- Random Forest (RF) is an extension of Bagging that builds multiple decision trees using random feature subsets to reduce overfitting, improve generalization, and handle high-

dimensional data efficiently.

- Support vector machine (SVM) is a supervised learning algorithm that classifies data by finding the hyperplane with the maximum margin between classes, using kernel functions to handle non-linearly separable data, thus enhancing generalization performance.
- XGBoost (XGB) is a gradient boosting algorithm optimized for speed and performance, known for high accuracy and advanced regularization to prevent overfitting, with parallel processing capabilities for large datasets.

B. Base Model Selection and Training

From the seven available algorithms (ADA, DT, KNN, LR, RF, SVM, XGB), six are chosen to serve as the base models. The selection process can be based on prior performance, diversity, or other relevant criteria. Each of the six base models is independently trained on the Training subset of the ECG dataset, which contains the normalized feature values and corresponding arrhythmia labels. This training aims to optimize each model's parameters to accurately classify four arrhythmia classes including AFIB, SB, SR, and GSVT.

After training, the six base models generate predictions on the Training and Testing subsets. These predictions can either be probability scores for each class or direct class labels, depending on the meta-model's requirements. The outputs from the base models on the Training set are used to create a new dataset, which serves as input for the meta-model. This new dataset essentially captures the decision patterns of the base models, providing a higher-level representation of the data.

C. Meta-Model Training and Validation

The meta-model, which is actually selected as the remaining algorithm from the set of basic algorithms after six algorithms have been chosen as base models, is trained using the outputs of the base models as input features. This meta-model is designed to learn how to optimally combine the predictions of the base models to improve the overall classification accuracy. The Training data for the meta-model consists of the base models' predictions from the Training subset, along with the true class labels.

Once the meta-model is trained, it is used to make final predictions on the Testing subset. The meta-model takes the outputs of the six base models on the Testing set as input and generates the final classification. A 3-fold cross-validation is used for the meta-model to ensure that the developed classifier learns a robust and generalized way to combine the predictions from the base models. By leveraging the strengths and mitigating the weaknesses of each base model, the meta-model aims to provide more accurate and robust arrhythmia classification results of four arrhythmia types.

D. Classification Model Development

All seven blending ensemble learning models, combinations of an individual ML algorithm as the meta-model and six others as base models, have been studied. The performance of the blending ensembles is evaluated using standard metrics such as accuracy, F1-score, precision, and recall. The most efficient blending ensemble solution for the classification of the four cardiac arrhythmia classes is chosen based on the superior performance exhibited by the proper blending ensemble models.

IV. Experimental Results and Discussion

A. Performance Metrics

Key performance metrics—Accuracy, F1-score, Precision, and Recall—are used to evaluate the machine learning models in the numerical experiments. Accuracy reflects the overall prediction correctness, Precision measures the correctness of positive predictions, Recall assesses the ability to identify all true positives, and the F1-score balances classification across the four arrhythmia labels. The evaluation includes overall accuracy, as well as macro, micro, and weighted averages of F1-score, Precision, and Recall: macro average treats each class equally, micro average aggregates metrics across all classes, and weighted average adjusts for class size to account for imbalances.

B. Base Models

All seven ML algorithms are considered as possible base models in blending ensemble models for 12-lead ECG-based classification of 4 arrhythmia types including AFIB, SB, SR, and GSVT. These base models are trained independently on the training data to generate the predicted probabilities, which are then used as input features for a meta-model, of the four arrhythmia classes, and the final performances on the testing set are evaluated. It is important to note that hyper-parameter tuning is performed using Grid Search to identify the optimal model and its best configuration. Table I summarizes the obtained results for all seven algorithms. The role of the base models is to capture diverse patterns and strengths from the data, and their combined outputs are utilized by the meta-model to learn how to optimally integrate these predictions, enhancing overall model performance.

TABLE I
ACCURACY COMPARISON

Classification		Base models							Meta models						
		ADA	*DT*	*KNN*	*LR*	*RF*	*SVM*	*XGB*	***ADA***	***DT***	***KNN***	***LR***	***RF***	***SVM***	***XGB***
Each class	***AFIB***	0.954	0.932	0.973	0.977	0.950	0.980	0.976	*0.977*	*0.975*	*0.977*	*0.979*	*0.977*	*0.977*	*0.974*
	SB	0.987	0.978	0.981	0.985	0.991	0.987	0.993	*0.993*	*0.992*	*0.993*	*0.994*	*0.993*	*0.993*	*0.991*
	SR	0.981	0.954	0.973	0.975	0.968	0.977	0.986	*0.986*	*0.983*	*0.986*	*0.986*	*0.985*	*0.986*	*0.986*
	GSVT	0.958	0.972	0.977	0.972	0.979	0.974	0.983	*0.984*	*0.983*	*0.985*	*0.985*	*0.983*	*0.984*	*0.977*
Overall		0.940	0.917	0.952	0.954	0.944	0.958	0.969	*0.970*	*0.966*	*0.970*	***0.972***	*0.968*	*0.970*	*0.964*

C. Performance of Blending Ensembles

Blending ensembles are constructed by combining 6 base models among the set of 7 fundamental algorithms and the remainder as the meta-model. Hence, there are seven blending ensembles with the meta-models of ADA, DT, KNN, LR, RF, SVM, and XGB respectively. For each blending ensemble, a meta-dataset is created by merging the predicted probabilities of the 6 other algorithms served as base models. The meta-model is then trained on the meta dataset along with the true labels from the training data. Its goal is to learn how to optimally combine the base models' predictions to improve overall performance. Once trained, the meta-model uses the base models' predictions on the testing set to make the final prediction. The meta-model's performance is then evaluated using appropriate metrics, ensuring that it effectively integrates the diverse outputs of the base models to achieve better accuracy, generalization, and robustness in predictions.

The performance comparisons of the blending ensembles with different meta-models of ADA, DT, KNN, LR, RF, SVM, and XGB, for arrhythmia classification across four classes: AFIB, SB, SR, and GSVT, in terms of accuracy and F1-score, are shown in Table I and Figure 2 correspondingly. It implies that Logistic Regression (LR) consistently achieves the highest performance, with an overall score of 0.972 and the highest accuracy in most individual classes, particularly AFIB (0.979) and SB (0.994). The performance across all meta-models is relatively similar, with only slight variations in accuracy, indicating that each algorithm captures the predictive patterns effectively. However, the DT and XGB models generally show slightly lower overall scores, suggesting they may be less optimal choices compared to others like LR or KNN. These results highlight the blending ensemble with the LR meta-model as a robust arrhythmia classification, providing balanced and high performance across all arrhythmia classes. The performance of the proposed blending ensemble with the LR meta-model is given in Table II. It is verified that the blending ensembles offer better performance compared to individual machine learning algorithms.

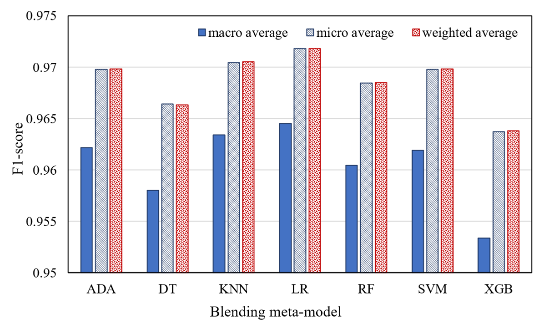

Fig. 2. F1-score comparison

TABLE II
BLENDING MODEL PERFORMANCE

Classification		Parameters			
		Acccuracy	*F1-score*	*Precision*	*Recall*
Individual class	*AFIB*	0.979	0.964	0.934	0.949
	SB	0.994	0.992	0.993	0.992
	SR	0.986	0.969	0.977	0.973
	GSVT	0.985	0.929	0.958	0.944
Average	*macro*	0.972	0.965	0.966	0.964
	micro		0.972	0.972	0.972
	weighted		0.972	0.972	0.972

TABLE III
PERFORMANCE COMPARISON

Performance parameters		Comparative methods		
		[15]	*[14]*	*This work*
Overall accuracy		0.97	0.842	0.972
F1-score	*AFIB*	0.941	0.83	0.964
	SB	0.993	0.889	0.992
	SR	0.977	0.964	0.969
	GSVT	0.949	0.897	0.929
	macro ave.	0.965	-	0.965
	micro ave.	0.970	-	0.972
	weighted ave.	0.970	-	0.972

D. Discussion

We developed an efficient blending ensemble learning solution integrating 6 base models (ADA, DT, KNN, RF, SVM, and XGB) with the LR meta-model for multi-classes arrhythmia classification using 12-lead ECGs. Without the necessity of the clinical features, i.e. age or gender, ..., our proposed solution outperforms conventional works given in

[15] and [14]. As shown in Table III, the proposed method achieves the highest overall accuracy of 0.972, slightly outperforming [15] (0.97) and significantly surpassing [14] (0.842). For individual F1-scores, the macro, micro, and weighted averages of the proposed method are all high (0.965, 0.972, 0.972 respectively), demonstrating that it provides a balanced and effective classification performance overall, making it a robust and competitive approach for arrhythmia classification. The reason behind the success of our method comes from the efficiency of the blending which is a powerful ensemble technique that can significantly boost the performance of machine learning models by leveraging the complementary strengths of diverse algorithms. It is important to mention that the way we set up the cross-validation procedure also contributes significantly to our final results.

V. Conclusion

Blending models represent a powerful yet accessible approach to ensemble learning, combining the predictive strengths of multiple models to achieve superior performance. In this paper, we have developed an efficient and robust machine learning solution for four-classes arrhythmia classification using 12-lead ECGs through a blending ensemble learning approach based on seven fundamental and diverse machine learning algorithms including ADA, DT, KNN, LR, RF, SVM, and XGB. All blending configurations integrating six base models and a meta-model among the seven ML algorithms have been investigated, and the most efficient blending ensemble solution using the LR meta-model has been verified and proposed. The proposed blending solution significantly enhances classification accuracy, achieving up to 97.2%, demonstrating superior performance compared to conventional methods. The use of a LR meta-model not only improves interpretability and generalization but also reduces the risk of overfitting, making the model a reliable tool for clinical decision support. These findings highlight the potential of blending ensemble learning to enhance automated arrhythmia classification, providing valuable insights that could inform future applications in cardiovascular disease management.

References

[1] W. H. Organization, "Cardiovascular diseases," https://www.who.int/europe/news-room/fact-sheets/item/cardiovascular-diseases, 2024, accessed: 15 August 2024.

[2] L. et. al., "A review of arrhythmia detection based on electrocardiogram with artificial intelligence," *Expert review of medical devices*, vol. 19, no. 7, pp. 549–560, 2022.

[3] S. et. al., "Machine learning approach to detect cardiac arrhythmias in ecg signals: A survey," *Irbm*, vol. 41, no. 4, pp. 185–194, 2020.

[4] E. Merdjanovska and A. Rashkovska, "A framework for comparative study of databases and computational methods for arrhythmia detection from single-lead ecg," *Scientific Reports*, vol. 13, no. 1, p. 11682, 2023.

[5] M. et. al., *Comprehensive electrocardiology*. Springer Science & Business Media, 2010.

[6] L. et. al., "Two-dimensional ecg-based cardiac arrhythmia classification using dse-resnet," *Scientific Reports*, vol. 12, no. 1, p. 14485, 2022.

[7] J. et. al., "Cardiologist-level interpretable knowledge-fused deep neural network for automatic arrhythmia diagnosis," *Communications Medicine*, vol. 4, no. 1, p. 31, 2024.

[8] Q. et. al., "Cardiac arrhythmias classification using photoplethysmography database," *Scientific Reports*, vol. 14, no. 1, p. 3355, 2024.

[9] D. et. al., "Computer-aided arrhythmia diagnosis with biosignal processing: A survey of trends and techniques," *ACM Computing Surveys (CSUR)*, vol. 52, no. 2, pp. 1–37, 2019.

[10] D. U. Jeong and K. M. Lim, "Convolutional neural network for classification of eight types of arrhythmia using 2d time–frequency feature map from standard 12-lead electrocardiogram," *Scientific reports*, vol. 11, no. 1, p. 20396, 2021.

[11] Z. et. al., "Optimal multi-stage arrhythmia classification approach," *Scientific reports*, vol. 10, no. 1, p. 2898, 2020.

[12] H. et. al., "Comparison of machine learning algorithms using manual/automated features on 12-lead signal electrocardiogram classification: A large cohort study on students aged between 6 to 18 years old," *Cardiovascular Engineering and Technology*, vol. 14, no. 6, pp. 786–800, 2023.

[13] A. R. et. al., "Convolutional neural networks for electrocardiogram classification," *Journal of Medical and Biological Engineering*, vol. 38, pp. 1014–1025, 2018.

[14] A. et. al., "Ecg-based machine-learning algorithms for heartbeat classification," *Scientific reports*, vol. 11, no. 1, p. 18738, 2021.

[15] Z. et. al., "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," *Scientific data*, vol. 7, no. 1, p. 48, 2020.

[16] K. Gautam, *Ensemble methods for machine learning*. Manning, 2023.

[17] Y. et. al., "Applications of stacking/blending ensemble learning approaches for evaluating flash flood susceptibility," *International Journal of Applied Earth Observation and Geoinformation*, vol. 112, p. 102932, 2022.

[18] D. M. et. al., "NeuroKit2: A python toolbox for neurophysiological signal processing," *Behavior Research Methods*, vol. 53, no. 4, pp. 1689–1696, feb 2021. [Online]. Available: https://doi.org/10.3758%2Fs13428-020-01516-y