Detecting Experts Using a MiniRocket: Gaze Direction Time Series Classification of Real-Life Experts Playing the Sustainable Port

Gianluca Guglielmo(✉), Michal Klincewicz, Elisabeth Huis in 't Veld, and Pieter Spronck

Tilburg University, Tilburg, Netherlands
G.Guglielmo@tilburguniversity.edu

Abstract. This study aimed to identify real-life experts working for a port authority and lay people (students) who played The Sustainable Port, a serious game aiming to simulate the dynamics occurring in a port area. To achieve this goal, we analyzed eye gaze data collected non-invasively using low-grade webcams from 28 participants working for the port authority of the Port of Rotterdam and 66 students. Such data were used for a classification task implemented using a MiniRocket classifier, an algorithm used for time-series classification. The classifier reached an F1 score of 0.75 (SD = 0.07), a PR AUC of 0.73 (SD = 0.14), and an ROC AUC of 0.75 (SD = 0.15) providing evidence that it is possible to identify real-life experts about maritime port management using data that can be obtained from a webcam. We speculate that the gaze direction used to train the MiniRocket may contain relevant information about the cognitive processes and decisions occurring throughout the gameplay. We suggest that the methods here presented not only can be used to detect experts playing simulations, such as serious games, but also to identify experts tackling screen-presented tasks.

Keywords: Gaze · Expertise · Machine Learning · Maritime Port · Serious Games · Sustainability · Time Series Classification

1 Introduction

Serious games are valuable tools that have already been used for training, educational purposes, and hiring new employees [6]. This last point seems to suggest that relevant skills that a candidate may have can emerge during the game play of a serious game; such skills may be those that experts in a specific field already have. Therefore, it may be not surprising that experts in real-life play and experience a serious game differently than lay people. One such example is The Sustainable Port [2], a serious game that simulates the real-world environment of the Port of Rotterdam (PoR). Players put themselves into the role of

A. Schönbohm et al. (Eds.): GALA 2024, LNCS 15348, pp. 177–187, 2025.
https://doi.org/10.1007/978-3-031-78269-5_17

executive decision-makers at PoR with the goal of balancing the port's profits and emissions. Seventy percent of PoR employees who played The Sustainable Port reported that they used their previous experience acquired at the Port of Rotterdam during their gameplay and indeed, they achieved better scores than university students [14]. The main goal of the current study is to evaluate whether it is possible to discriminate experts from lay people based on their eye movements during the gameplay using non-invasive methods such as the eye gaze information extracted from OpenFace [3] combined with a fast algorithm for timeseries classification such as the MiniRocket [9]. The secondary goal was to further test the assumption that real-life expertise is applied in a simulation and that this transfer is detectable not only by overt, conscious actions and decisions during the gameplay, but also by covert, automatic physiological changes tracked by non-invasive methods. The results of this study explore additional avenues in which serious games, combined with accessible technology, can detect interesting decision-making profiles for hiring or training purposes.

2 Expertise, Physiology, and Gaze

Prior experience and knowledge of experts affects their behavior and physiological responses. Experts show different eye-related patterns during a task compared to their less experienced counterparts. For example, expert radiologists have different eye gaze patterns of fixation and saccades while looking at X-rays [25]. One way to interpret this pattern of results is that eye gaze [12] reflects more effective working memory use, with experts being more effective in encoding and retrieving task-related information [10]. Another possible explanation is that experts are more effective in the selective allocation of their attentive resources focusing on relevant stimuli and neglecting the non-relevant ones [12,15]. What may also play a role, at least for some of these differences, is familiarity with the task [2], which is reflected in more efficient use of working memory or more effective allocation of attentive resources [5].

3 Methods

3.1 The Sustainable Port Game

The Sustainable Port is a game simulating port dynamics introduced by a previous study by Guglielmo, Klincewicz, Huis in 't Veld, and Spronck [14]. Such a game unfolds across ten rounds, and the player has to adjust their strategy across the game to balance CO2 emission and the Added Value (the revenues). The final score obtained in the game is calculated by subtracting the CO2 emissions from the Added Value. Furthermore, all players know that they need to have a CO2 $<=$ 10 in order not to lose the game outright and that their score at the end of round 10 is the final measure of performance. During the game, the participants can build, destroy, upgrade, and close facilities.

3.2 Sample

In this experiment, a total of N = 109 people were recruited to play The Sustainable Port, of which n = 75 students (Nmales = 27, Nfemales = 47 females, Nnd = 1; Mage = 21.20, SD = 3.4) from Tilburg University and n = 34 (Nmales = 23, Nfemales = 11; Mage = 40.67, SD = 10.82) from the Port of Authority of the Port of Rotterdam. Two PoR employees were excluded from the analysis due to a corrupted video recording. Additionally, 4 PoR employees and 9 students were excluded due to their inability to understand the game mechanics. More specifically, the participants were asked the following question: "During which round did you develop confidence in the game mechanics? (for example: which button or option is associated to specific actions)". The final sample used for analysis consisted of N = 94 participants (28 PoR employees and 66 students).

3.3 Data Collection Procedure

The students were recruited through the recruitment system at Tilburg University in exchange for a formative credit. PoR employees were recruited from the HRM department, Strategy department, Finance department, Environmental management, Port Development department, Commercial Department & Policy Department of the Port Harbour Master. All of these departments were relevant for our study since they are involved in strategic decisions concerning the green transition and the future development of the Port of Rotterdam. Participants were recruited anonymously with the help of a spreadsheet shared through the internal Port of Rotterdam newsletter. Interested employees were asked to create an alphanumeric code making sure they do not use any information about their identity (like their date of birth or name). Once they signed up, all participants were informed that the study was approved by the ethics committee of Tilburg University (REDC 2021.35f + study number 1) and that it would involve expertise and decision-making.

Both the participants' faces and the gameplay were recorded using the Open Broadcaster Software [18]. After the final 10th round, participants were presented with their score and whether they reached the necessary CO2 threshold to not lose the game. They were also asked, through a questionnaire, to indicate during which round they understood the game mechanics (or to declare that they did not understand them at all) if they paid attention to the game dynamics, and if they were motivated to play the game on a Likert scale between 1 (strongly disagree) to 7 (strongly agree), Each experimental session lasted approximately 60 min.

3.4 Preprocessing

To extract gaze direction, the video recordings were cropped to just contain the faces of participants. Gaze direction was extracted, using OpenFace [3], from eye gaze angle as two series of values: `gaze_angle_x` and `gaze_angle_y`, which respectively refer to the horizontal axis and the vertical axis of the area that participants are looking at during the gameplay.

Participants took 19.90 min (SD = 6.63) on average to finish all 10 rounds of the game. All the recordings had a variable length with PoR employees spending significantly more time playing the game (M = 23.88, SD = 6.48) than students (M = 18.21, SD = 5.94; Welch t-test results: $t(46.82) = 3.92$, $p < .001$). Therefore, the data extracted from the video recordings were padded using the single individual mean of each participant for the `gaze_angle_x` (horizontal movements) and the `gaze_angle_y` (vertical movements), so that they had the same lengths [4]. In this specific case, we used the single individual mean of each participant for the `gaze_angle_x` (horizontal movements) and the `gaze_angle_y` (vertical movements). This approach has several advantages over other methods. First, it is better than truncation given evidence of its non-robustness [4] and the distortion it generates to the original time series. Second, zero padding, a commonly used technique, was also excluded because it may introduce a bias. This 0-based bias would likely arise due to students and PoR employees having significantly different lengths of average recordings. A significantly higher number of zeros at the end of the recordings for a specific group (as it would occur for the students' group), due to padding, may result in potentially extracting misleading features from such zeros. Given these reasons, we chose the single mean padding [23], which is based on the mean of the distribution of gaze direction for each participant and should not introduce class-specific bias. The padded-time series used as inputs were as long as the longest time series in our data (69361 data points which corresponds to 38.53 min at 30 fps). At the end of the classification task, we still compared the padded length of the misclassified instances with the padded length of the correctly classified instances. This approach was used to evaluate if the padded length played a role in correctly classifying or misclassifying our participants even after using the individual means of each participant. If the padded length played a role in correctly identifying the two groups we would expect the correctly classified instances and the incorrectly classified ones to have a significant difference in their padded length. Such analyses were run after controlling for the normality of the residuals with a KS test [27] and the homogeneity of variance with the Bartlett test [32]. Besides evaluating the effect of padding, we also trained the same classifier (MiniRocket), using the truncated version of the time series. Therefore, all the time series were truncated at 15902 datapoints (8.83 min) corresponding to the shortest recording in our data. Such a process was adopted to evaluate if even in case of reduced and limited information our classifier still performs above baseline suggesting emerging differences between the two groups before the end of the recordings.

3.5 Evaluation of Confounding Variables

Even though previous studies found a significant correlation between the ranges of gaze angles and age due to younger individuals having presumably wider eye movements [21] no significant correlation (Pearson's r coefficient) was found between age and vertical gaze ($r = -0.07$, $p = .51$) or horizontal gaze ($r = -0.09$, p = .38) in our sample. Furthermore, given that previous studies showed that attention [28] and motivation [17] influence eye gaze, we evaluated if there was a significant difference between the two groups we analyzed. Our results show no significant difference in self-assessed motivation (PoR employees M = 5.64, SD = 0.97; Students M = 5.55, students SD = 1.16; t(59.60) = 0.41, p = .68) and self-assessed attention (t(42.48) = 0.14, p = .89; PoR employees M = 5.54, PoR employees SD = 1.18; Students M = 5.50, Students SD = 0.96) between the two groups.

3.6 MiniRocket Time Series Classifier

MiniRocket (MINImally RandOm Convolutional KErnel Transform) is an algorithm for time series classification. Compared to other time series classification algorithms, it offers an optimal balance between accuracy, computational power, and computational complexity [7,9]. MiniRocket has two uniquely engineered phases: a feature extraction phase and a classification phase. Features are extracted using randomly generated convolutional kernels with a fixed length of 9 [7]. The kernel operation in the MiniRocket maps a problem that is not linearly separable, as in most time series problems, into a multi-dimensional space that makes them linearly separable [20]. Like convolutional neural networks, zero padding is applied at the beginning and at the end of the input, so that convolution operations are centered on values composing it. The weights are drawn between -1 and 2 while the bias is directly obtained from the convolution output [9]. After that, the generated kernels are convoluted with the input data (the time series used as input for the MiniRocket) to generate the final set of features that should make the problem linearly separable.

A single feature per kernel is extracted: the proportion of positive values pooling (PPV), which is the output of each convolution operation. The PPV is the proportion of positive values in the convoluted region [9]. This process retains the most essential information reducing computational complexity [19]. However, unlike the convolutional neural network, in the MiniRocket there are no hidden layers and no linear combinations of features [9]. MiniRocket does not have many hyperparameters and they are not supposed to be tuned, such hyperparameters are the max dilations per kernel (equal to 32) and the number of kernels, which is 9996. From each kernel, the PPV is extracted thereby always generating 9996 features (the number of kernels multiplied by the PPV extracted). After the features are extracted, during the second phase (the classification phase), these features are used to train a linear classifier, such as a Support Vector Machine, a Logistic Regression, or in the case of small datasets a Ridge Classifier [9].

MiniRocket proved to have satisfactory performance on several datasets compared to competitors' algorithms [9] while maintaining a relatively small computation time [1]. Furthermore, this algorithm has already been successfully used for different purposes: from virtual price currency prediction [7] to identifying subjects with atrial fibrillation [1]. However, despite its success, one of the problems of the MiniRocket is potential multicollinearity among extracted features (the PPVs) which can contain noise generating significant correlation among the features [20]. This can be mitigated by using a Ridge Classifier, with an optimized alpha shrinkage parameter, known to be resistant to multicollinearity [8].

3.7 Pipeline for Classification Task and Model Evaluation

The MiniRocket was combined with a ridge classifier using sktime given our small sample. We also used the default options for in MiniRocket in sktime, and we did not scale the time series inputs since MiniRocket does not require this process [9]. This approach was demonstrated effective for small datasets [1] and on multivariate time series data [29]. A 5-fold stratified cross-validation respecting the original class imbalance present in the dataset (approximately composed of 70% students and 30% PoR employees) was used. During each of the 5 folds, the data were partitioned so that 80% of the data was used as a train set and 20% as a test set. During each fold, the train set is split again using another 5-fold stratified cross-validation to detect the best alpha parameter for the ROC AUC, given its alleged robustness towards class imbalances [30]. The range of values for alpha used for the optimization was between 10 and 100 opting for a stronger regularization to better cope with potential multicollinearity and overfitting considering the small size of our dataset. The input for the MiniRocket consists of a dataset of 94 participants (instances) with 2 dimensions (`gaze_angle_x` and `gaze_angle_y`) where each time series had a length of 69361 data points. This 2-dimensional input data (the mean padded time series of the `gaze_angle_x` and `gaze_angle_y`) was fed to the MiniRocket algorithm that extracted the PPV value for each of the 9996 kernels. After having extracted the features (the PPV extracted per kernel), the RidgeClassifierCV function available in sklearn was used for the classification task, balancing the class weights during training.

The ROC, AUC, weighted F1 score, and PR AUC metrics were used to evaluate the performance of the classifier (considering the PoR employees as the positive class) [16,22,30]. The extracted sample weights for the train splits (during each fold) were applied when computing the PR AUC; this process was adopted since sklearn does not have a weighted option for the PR AUC (examples of how to calculate the PR AUC with sklearn can be found at the following link: https://sinyi-chou.github.io/python-sklearn-precision-recall/). The evaluation of the model was done after the optimal value of alpha for the ROC AUC was detected on each of the 5 train splits generated by the 5-fold stratified cross-validations. The final metrics (ROC AUC, PR AUC, and F1) were first calculated on the test set for each original split and after that, at the end of the process, averaged across all the 5-folds defined. The same process, with the same data split, was used to train another MiniRocket classifier with the truncated version of the time series (94 participants with 2 dimensions, `gaze_angle_x` and `gaze_angle_y`, where each time series had a length of 15902 data points in this case).

Finally, we compared MiniRocket's performance with a Dummy Classifier that uses a "constant" as strategy to find the baseline F1 and "stratified" as strategy to find the baseline ROC AUC and PR AUC. The Dummy Classifier, used to define classification baselines, ignores the features provided and just looks at the distributions of the instances according to the class labels (More information about the Dummy Classifier and its classification strategies can be found at: https://scikit-learn.org/stable/modules/model_evaluation.html).

4 Results

MiniRocket managed to correctly identify around 70% of the PoR employees and 74% of the students. After having extracted a confusion matrix, we found that, across the entire dataset on the test splits of the 5-fold stratified cross-validation, MiniRocket correctly detected 19 PoR employees (9 misclassified as students) and 51 students (15 misclassified as PoR employees). Overall, MiniRocket correctly detected 70 participants out of the 94 collected. Furthermore, the classifier always outperformed the Dummy Classifier (baseline) as displayed in Table 1.

Table 1. The scores of the chosen metrics and the comparison with the Dummy Classifier (baseline).

	F1	ROC AUC	PR AUC
MiniRocket	0.75 (SD = 0.07)	0.75 (SD = 0.14)	0.73 (SD = 0.15)
Dummy Classifier	0.46 (SD = 0.03)	0.70 (SD = 0.00)	0.45 (SD = 0.13)

The padded lengths of the correctly and incorrectly classified instances did not differ for either the PoR employees (Correct: M = 23.95 min, SD = 2.90. Incorrect M = 19.90 min, SD = 2.90; $t(35) = 1.77$, $p = .09$), or the students (Correct: M = 18.32 min, SD = 4.60. Incorrect M = 21.37 min, SD = 5.77; $t(24.48) = 1.93$, $p = .08$). Furthermore, when evaluating the results obtained by the MiniRocket using the truncated data we found out that even in this case the classifier performed above baseline (Dummy Classifier). The MiniRocket with the truncated data obtained a mean F1 score of 0.66 (SD = 0.12), a mean ROC AUC of 0.59 (SD = 0.14), and a mean PR AUC = 0.56 (SD = 0.21). In this case, using truncated data, 10 PoR employees (18 misclassified) and 51 students (15 misclassified) were successfully detected across the 5-fold stratified cross-validation.

5 Discussions

The results here presented suggest that PoR employees can be discriminated from lay people (students) using information contained in gaze direction. MiniRocket, combined with OpenFace gaze information [3] could therefore offer a fast, effective, and accessible solution to detect real-life experts playing serious games. Our results show the classifier likely does not rely on group differences in age [21], self-assessed motivation [28], or attention during gameplay [17]. Furthermore, we saw that padding seems to have a limited role or no role at all in our classification task; this is further confirmed when looking at the results obtained using truncated data. Such results, despite not reaching an acceptable ROC AUC of 0.70 [26], still obtained a performance above baseline in all the metrics further suggesting the limited effect of padding and the emergence of differences even before the classifier was provided with complete information about the recordings. Altogether, these results suggest that the classifier was actually able to successfully detect participants in the two groups and the effect of the considered confounders was lacking or negligible.

Unfortunately, the features used for the classification task by the MiniRocket have no intuitive interpretation being based on kernels [33]. However, it is likely that the PoR employees used what they already know about port environments to play the game better than students [14] as the game introduces very niche innovative technologies such as hydrogen-oriented technology [31], which is very related to the expertise of the PoR employees involved in this study, who all work for departments that actively participate in the modern debate about green transition. Furthermore, the gaze direction observed in these two groups may reflect different ways of dealing with the presented information, which may be connected to experiences obtained in similar informational environments [10] or different styles of suppressing irrelevant information [2]. If that is right, then our study provides some evidence that expertise learned in real life is reflected in behavior and physiological responses occurring during serious games. Another possibility is that differences in gaze patterns reflect different decisions made during gameplay. PoR employees performed better than students [14], so what we may be seeing in gaze could be connected to the specific in-game mechanics, and actions that experts happen to do during play. We can assume that similar behaviors may emerge in other serious games and screen-presented tasks. Therefore, future studies may focus on using these methods with other serious games and screen-presented tasks to evaluate their degree of generalizability. Obtaining similar results may pave the way for using the method here proposed for business applications to train or hire new employees [6].

Despite the promising results, our study has several limitations. First, our dataset was small and strongly imbalanced (notably, this reflects the distribution of experts in the general population) and we just presented results from one game: The Sustainable Port. Furthermore, besides the results obtained by the MiniRocket, there are still problems that need to be addressed when it comes to using this classifier. MiniRocket features are based on kernels that are not humanly interpretable [33] even though methods such as SHAP could help detect the relevant kernels used for classification [24]. However, a more transparent algorithm would yield better explanations of performance. Future studies should apply other, more computationally expensive and time-consuming, classifiers such as learning shapelets [31] or pattern-based embedding for time series classification [13] to have more accessible interpretations of the results. Third, using other sources of behavioral and physiological signals, such as the ones extractable from eye trackers [11] may provide better and more information to discriminate experts from lay people or novices [34] compared to the gaze angle extracted in OpenFace; however, such data may require more expensive tools to be collected. If the results we report here can be replicated in other serious games and if a more interpretable classifier combined with eye-tracking data is used with them, we may obtain deeper insights into the physiological patterns and behaviors making specific people better decision-makers in complex environments.

Acknowledgments. This research is founded by the MasterMinds project, part of the RegionDeal Mid- and West-Brabant, and it is co-founded by the Ministry of Economic Affairs and Municipality of Tilburg.

References

1. Alagoz, C.: Robust and efficient atrial fibrillation detection from intracardiac electrograms using MiniRocket. Int. J. Eng. Res. Dev. **16**(1), 432–447 (2024)
2. Alonso, A., van der Meij, J., Tse, D., Genzel, L.: Naïve to expert: considering the role of previous knowledge in memory. Brain Neurosci. Adv. **4**, 2398212820948686 (2020)
3. Baltrusaitis, T., Zadeh, A., Lim, Y., Morency, L.: OpenFace 2.0: facial behavior analysis toolkit. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018) (2018)
4. Bier, A., Jastrzębska, A., Olszewski, P.: Variable-length multivariate time series classification using rocket: a case study of incident detection. IEEE Access **10**, 95701–95715 (2022)
5. Brams, S., et al.: The relationship between gaze behavior, expertise, and performance: a systematic review. Psychol. Bull. **145**(10), 980–1027 (2019)
6. Bylieva, D., Lobatyuk, V., Rubtsova, A.: Serious games as a recruitment tool in educational projects. In: European Proceedings of Social and Behavioural Sciences, vol. 51 (2018)
7. Byrd, J., Lin, B., Loesch, F., Neubarth, M., Peng, Z., Tian, F.: The application of minirocket in virtual currency price prediction (2024)
8. Cule, E., Iorio, M.D.: Ridge regression in prediction problems: automatic choice of the ridge parameter. Genet. Epidemiol. **37**(7), 704–714 (2013)
9. Dempster, A., Schmidt, D., Webb, G.: MiniRocket: a very fast (almost) deterministic transform for time series classification. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, pp. 248–257 (2021)
10. Ericsson, K., Kintsch, W.: Long-term working memory. Psychol. Rev. **102**, 211–245 (1995)
11. Feremans, L., Cule, B., Goethals, B.: PETSC: pattern-based embedding for time series classification. Data Min. Knowl. Disc. **36**(3), 1015–1061 (2022)
12. Gegenfurtner, A., Lehtinen, E., Säljö, R.: Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains. Educ. Psychol. Rev. **23**, 523–552 (2011)

13. Grabocka, J., Schilling, N., Wistuba, M., Schmidt-Thieme, L.: Learning time-series shapelets. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 392–401 (2014)
14. Guglielmo, G., Klincewicz, M., Huis in 'T. Veld, E., Spronck, P.: Introducing sustainable port. In: Proceedings of the International Conference of Games, Entertainment, and Media (GEM). IEEE (2024)
15. Haider, A., Frensch, P.: Eye movement during skill acquisition: more evidence for the information-reduction hypothesis. J. Exp. Psychol. Learn. Mem. Cogn. **25**, 172 (1999)
16. Jeni, L., Cohn, J., Torre, F.D.L.: Facing imbalanced data–recommendations for the use of performance metrics. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, pp. 245–251 (2013)
17. Khan, A., Blohm, G., McPeek, R., Lefevre, P.: Differential influence of attention on gaze and head movements. J. Neurophysiol. **101**(1), 198–206 (2009)
18. Kristandl, G.: "All the world's a stage"–the open broadcaster software (OBS) as enabling technology to overcome restrictions in online teaching. Compass: J. Learn. Teach. Univ. Greenwich **14**(2), 1–16 (2021)
19. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, vol. 25 (2012)
20. Kung, S.: Kernel Methods and Machine Learning. Cambridge University Press, Cambridge (2014)
21. Lee, W., Kim, J., Shin, Y., Hwang, S., Lim, H.: Differences in eye movement range based on age and gaze direction. Eye **33**(7), 1145–1151 (2019)
22. Lu, H., Ehwerhemuepha, L., Rakovski, C.: A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance. BMC Med. Res. Methodol. **22**(1), 181 (2022)
23. Mann, M.: Smoothing of climate time series revisited. Geophys. Res. Lett. **35**(16) (2008)
24. Marcílio, W.E., Eler, D.M.: From explanations to feature selection: assessing SHAP values as feature selection mechanism. In: 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), pp. 340–347. IEEE (2020)
25. McLaughlin, L., Bond, R., Hughes, R., McConnell, J., McFadden, S.: Computing eye gaze metrics for the automatic assessment of radiographer performance during X-ray image interpretation. Int. J. Med. Informatics **105**, 11–21 (2017)
26. Metz, C.: Basic principles of ROC analysis. Semin. Nucl. Med. **8**(4), 283–298 (1978)
27. Mishra, P., Pandey, C.M., Singh, U., Gupta, A., Sahu, C., Keshri, A.: Descriptive statistics and normality tests for statistical data. Ann. Card. Anaesth. **22**(1), 67 (2019). https://doi.org/10.4103/aca.ACA_157_18
28. Nikitin, J., Freund, A.: Age and motivation predict gaze behavior for facial expressions. Psychol. Aging **26**(3), 695 (2011)
29. Pantiskas, L., Verstoep, K., Hoogendoorn, M., Bal, H.: Taking rocket on an efficiency mission: a distributed solution for fast and accurate multivariate time series classification. In: Proceedings of the XYZ Conference, pp. 123–130 (2021)
30. Richardson, E., Smith, M., Doe, J., Johnson, A., Williams, L.: The ROC-AUC accurately assesses imbalanced datasets. Available at SSRN 4655233 (2024)
31. Sazali, N.: Emerging technologies by hydrogen: a review. Int. J. Hydrogen Energy **45**(38), 18753–18771 (2020)
32. Viwatwongkasem, C.: A comparison of type I error and power of Bartlett's test, Levene's test and Cochran's test under violation of assumptions. Ph.D. thesis (2004)

33. Wu, C., Miller, J., Chang, Y., Sznaier, M., Dy, J.: Solving interpretable kernel dimension reduction. arXiv preprint arXiv:1909.03093 (2019)
34. Zhu, M., Bao, D., Yu, Y., Shen, D., Yi, M.: Differences in thinking flexibility between novices and experts based on eye tracking. PLoS ONE **17**(6), e0269363 (2022)