



AKADÉMIAI KIADÓ

Journal of Behavioral Addictions

DOI:
10.1556/2006.2024.00063
© 2024 The Author(s)

COMMENTARY



Machine Learning(s) in gaming disorder through the user-avatar bond: A step towards conceptual and methodological clarity

Reply to: User-avatar bond as diagnostic indicator for gaming disorder: A word on the side of caution (2024)

VASILEIOS STAVROPOULOS¹ , MARIA PROKOFIEVA² , DANIEL ZARATE^{1*} , MICHELLE COLDER CARRAS³ , RABINDRA RATAN⁴ , RACHEL KOWERT^{5,6} , BRUNO SCHIVINSKI⁷ , TYRONE L. BURLEIGH¹ , DYLAN POULUS^{8,9} , LEILA KARIMI¹ , ANGELA GORMAN-ALESI¹⁰, TAYLOR BROWN¹ , RAPSON GOMEZ¹ , KAIDEN HEIN¹ , NALIN ARACHCHILAGE¹¹ and MARK D. GRIFFITHS¹²

¹ Department of Psychology, Applied Health, School of Health and Biomedical Sciences, RMIT University, Australia

² Victoria University, Australia

³ Bloomberg School of Public Health, John Hopkins University, Baltimore, USA

⁴ Michigan State University, USA

⁵ Psychology Research Institute, Faculty of Social Studies, Masaryk University, Brno, Czech Republic

⁶ Department of Communication, University of Münster, Münster, Germany

⁷ School of Media and Communication, RMIT University, Australia

⁸ Physical Activity, Sport and Exercise Research Theme, Faculty of Health, Southern Cross University, Gold Coast, Australia

⁹ Manna Institute, Southern Cross University, Gold Coast, Australia

¹⁰ Catholic Care Victoria, Australia

¹¹ School of Computing, RMIT University, Australia

¹² International Gaming Research Unit, Psychology Dept, Nottingham Trent University, UK

Received: August 12, 2024 • Revised manuscript received: September 30, 2024 • Accepted: October 12, 2024

ABSTRACT

In response to our study, the commentary by Infanti et al. (2024) raised critical points regarding (i) the conceptualization and utility of the user-avatar bond in addressing gaming disorder (GD) risk, and (ii) the optimization of supervised machine learning techniques applied to assess GD risk. To advance the scientific dialogue and progress in these areas, the present paper aims to: (i) enhance the clarity and understanding of the concepts of the avatar, the user-avatar bond, and the digital phenotype concerning gaming disorder (GD) within the broader field of behavioral addictions, and (ii) comparatively assess how the user-avatar bond (UAB) may predict GD risk, by both removing data augmentation before the data split and by implementing alternative data imbalance treatment approaches in programming.

*Corresponding author.
E-mail: daniel.zarate.psychology@gmail.com

KEYWORDS

machine learning, gaming disorder, user-avatar bond, classification, diagnosis



INTRODUCTION

In response to our paper (Stavropoulos et al., 2023), the recent commentary by Infanti et al. (2024) highlighted areas requiring further clarification regarding (i) the conceptualization and utility of the user-avatar bond (UAB) as a gaming disorder (GD) risk indicator, and (ii) the optimal application of supervised Machine Learning (ML) techniques in addressing prediction challenges related to imbalanced data, particularly within the context of the UAB-GD association.

Infanti et al. noted that: (i) the UAB may not apply to all game genres and differs across individuals, and (ii) digital phenotyping should involve superior objective data (e.g., screen/application monitoring). They also had concerns regarding methodology, including the (iii) use of synthetic minority oversampling technique (SMOTE) in algorithm development, (iv) the non-use of ML regression methods, (v) how the Gaming Disorder Test (GDT-4) was used, and (vi) the lack of access to the original data.

Such considerations are valuable to promote and foster dialogue and advance clarity in the field. To contribute to this discourse, the present commentary aims to: (i) further clarify the concepts of avatar, UAB, and digital phenotype in relation to GD within the broader context of behavioral addictions, and (ii) provide evidence regarding the UAB-GD-risk association by reproducing our original analysis without augmenting data before the data split, as well as indicatively using a combination of alternative data-balancing techniques.

CONCEPTUAL CLARITY

Infanti et al. (2024) suggest that data used in digital phenotypes should be “superior to self-report,” citing a GD digital phenotyping study by Montag and Rumpf (2021). However, in their study, Montag and Rumpf (2021) argued that “the most accurate picture” of internet use disorders can be found by “asking participants about symptom load related to IUD [internet use disorder]” (i.e., self-report data) in combination with objective recording. Because mental health conditions (including GD) involve disturbances in behaviors, cognitions, and emotions (American Psychiatric Association [APA], 2022), self-reports of thoughts and feelings, including those related to avatars, have always held an important role in diagnosis, with clinical interviews as the gold standard. As stated throughout our original paper, the intent of our study was to improve the ability to identify possible GD risk, not to substitute a comprehensive diagnostic procedure. Using ML to classify individuals as having GD risk (as in the use of ML for determining risk for any condition) based on the intensity of their GD symptoms and the UAB does not imply diagnosis, but rather assigning them to a high-risk category. From an epidemiological perspective, screening for the risk of mental health disorders is important for identifying those who may need treatment or prevention interventions based on their self-reported symptom levels (Eaton et al., 2012).

In that context, revisiting the origins and evolution of the digital phenotype concept can further support its relevance to UAB. Traditionally, a phenotype includes behaviors shaped by a person’s predispositions and life experiences, offering essential insights into their physical and mental health (Zarate, Stavropoulos, Ball, De Sena Collier, & Jacobson, 2022). The extended phenotype of an organism refers to all the environmental changes it generates, affecting its Darwinian fitness, or the possibility of its genes being passed to the next generation (i.e., surviving, reproducing; Dawkins, 1982). The digital phenotype, derived from the extended phenotype, refers to the transformations and traces that users leave in their online environment, captured by various detectors to achieve their species-specific aims (e.g., cultural recognition and acceptance; Loi, 2019).

As with other extended phenotypes, the interaction between users and their digital environment introduces a co-evolutionary bond (i.e., users change/affect their digital environment, which later changes/affects them in a perpetual spiral; Loi, 2019). Such observations prompted Haraway (2008) to metaphorically describe digital data as a companion species. Indeed, several studies have explored the phenotypical potential embedded within the use of online/digital media through passive ways (i.e., digital sensors) and/or active ways (e.g., self-report questionnaires in the context of digitally facilitated ecological momentary assessment [EMA]) under the digital phenotype umbrella term (Zarate et al., 2022). To enhance clarity, Zarate et al. (2022) proposed features of objectivity and granularity as key aspects of the digital phenotype, encompassing passive physiological sensing, digital biomarkers, mobile sensing, and cyber-phenotype (i.e., exclusively cyber-behavior) subtypes.

To acknowledge the species-specific functions (e.g., acceptance) and the dynamically co-evolving nature of the user-UAB association, the term ‘digital phenotype’ was introduced in quotation marks in our study (Stavropoulos et al., 2023). The choice of using digital phenotype as a conceptual proxy was reinforced by: (i) empirical evidence underscoring the significance of the phenotypical information conveyed by the way excessive users experience their connection with their avatar (e.g., Casale, Musicò, Gualtieri, & Fioravanti, 2023; Servidio, Griffiths, Boca, & Demetrovics, 2023; Szolin, 2022), and (ii) clinical evidence leveraging the insights included in the UAB to address disordered gaming use (e.g., Tisseron, 2009).

Infanti et al. (2024) additionally advocated that the UAB may not be operative as a GD risk indicator because avatars are not universally present across all game genres, and even when they are, the customization options and user experiences can vary significantly. This concern is understandable, as what an avatar constitutes is not consistently described in the literature, with more rigid definitions assuming a visual, often anthropomorphic depiction, whereas broader and more flexible conceptualizations include any type of representation (e.g., sounds, usernames, text-descriptions; “any representation of any controller” [Nowak & Fox, 2018; p. 34]). However, despite such discrepancies, there is a consensus that an avatar represents



the user digitally, even if not visualized, allowing them to interact with their digital environment and others within it (Nowak & Fox, 2018). In the end, representation and agency are the two necessary and distinctive avatar components (Nowak & Fox, 2018). Therefore, avatars and the bonds that users form with them, are present in every videogame in which the user's actions (agency) are reflected in some perceivable change within the game, and where a form of user portrayal is required (representation). Therefore, avatars can provide valuable insights for the user through projective processes, even if they are not anthropomorphic or embodied (Nowak & Fox, 2018; Tisseron, 2009). For instance, in a real-time strategy game, the player aims to expand their territory and influence using their in-game army, while in their offline life, they may experience a lack of space and control.

Additionally, Infanti et al. (2024) suggested that due to the variability in users' experience with their avatars, a fused relationship between the two may not always be present, complicating the inference of GD risk. We argue that it is exactly this variability of the UAB intensity, assumed to be normally distributed among the gamer population, irrespective of how this may be conceptualized (e.g., Banks & Bowman, 2021; Ratan & Dawson, 2016; Stavropoulos, Motti-Stefanidi, & Griffiths, 2022) that constitutes a source of information itself when dimensionally assessed. For example, users not at risk for GD may report lower levels of UAB, as observed in our study (Stavropoulos et al., 2023). Nevertheless, in our study, role-playing games were emphasized due to their inherent role-playing-avatar features and their higher association with GD risk (Mancini, Imperato, & Sibilla, 2019; Stavropoulos, Motti-Stefanidi, & Griffiths, 2022).

METHODOLOGICAL CLARITY

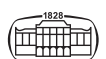
From a methodological perspective, Infanti et al. (2024) noted that supervised ML in our study may be affected by using the Synthetic Minority Oversampling Technique (SMOTE) before splitting the data into training and testing sets, and likely inflating an otherwise non-significant relationship. Although this approach has been applied by other ML health studies (e.g., Ishaq et al., 2021), to address this concern, while concurrently expanding on programming capabilities we (i) repeated the initial ML, R-based, analysis involving Naive Bayes, Random Forests, LASSO regression, logistic regression, k Nearest Neighbor (kNN), Support Vector Machine (SVM), and XG Boost without SMOTE before the data split, and (ii) employed, alongside SMOTE, three alternative data-imbalance remedies while training the MLs. These involved the Adaptive Synthetic Sampling Approach for Imbalanced Learning (ADASYN), the Random Over-Sampling Examples (ROSE), and the Tomek-Links approach (Khairy, Mahmoud, & Abd-El-Hafeez, 2024). ADASYN builds on SMOTE by creating synthetic samples while adapting their final number for each minority class instance based on the local density of the majority class

(He, Bai, Garcia, & Li, 2008). The ROSE algorithm performs over-sampling combined with smoothing in a multivariate way, and the Tomek-Links approach identifies case pairs from the majority and minority classes that are closest to each other representing borderline/noisy points, which are then used to undersample (Dar & Farooq, 2024; Haixiang et al., 2017). See Appendix for more information.

The random forests model using SMOTE post-split demonstrated acceptable performance (and superior to other MLs examined) considering all evaluated fit criteria across the three alternative data balancing approaches tested in predicting concurrent GD Risk. The findings achieved a Receiver Operating Characteristic (ROC) of 0.707, a Positive Predictive Value (PPV) of 0.808 (i.e., precision, how many of those classified as positive are truly positive), a Negative Predictive Value of 0.500 (NPV; i.e. how likely a person with a not at risk result truly is not at risk), a True Negative Rate of 0.934 (TNR; i.e. specificity or selectivity; how well a model identifies negative cases), a sensitivity of 0.877 (i.e. how well a test can identify true cases), an F-measure of 0.884 (F-meas; i.e., the ratio of the multiplication of recall and precision, multiplied by two, and then divided by the accumulation of recall and precision), a recall of 0.977 (i.e., true positive classified cases divided by the number of true positive cases), and an accuracy of 0.796 (i.e., the ratio of correctly predicted cases, across the total number of cases) (see Appendix). Considering predictions of prospective GD risk (i.e., six months later), the Random Forest model with ADASYN had collectively more balanced performance, comparatively to all other combinations of models and data-balancing modalities assessed across the different fit indices (ROC = 0.631, PPV = 0.808, NPV = 0.286, TNR = 0.167, F-meas = 0.848, Sensitivity = 0.894, recall = 0.894, and accuracy = 0.746). At this point, we wish to note that no other advanced combination of remedies (e.g., concurrent use of SMOTE and Tomek Links in conjunction with ensemble ML modeling, where a sequence of varying models are assembled to inform predictions) has been employed, as is beyond the scope of this commentary.

Infanti et al. (2024) criticized the non-use of ML regression methods, suggesting that more variable-focused approaches would better suit the assessment of GD risk. The classification approach employed was chosen deliberately based on the research aim. Rather than taking a variable-focused approach that identifies risk along a continuous spectrum, the goal was person-focused. It was to specifically highlight those individuals at GD risk. This approach is more practically meaningful, as it directs attention to those who may need intervention, rather than solely analyzing the overall severity of symptoms across a population.

Infanti et al. (2024) additionally highlighted that the four-item GDT-4 is intended to assess GD behavior severity, while the GDT-4 "functional impairment criterion" was not considered (WHO, 2022). It is important to reiterate that our use of the GDT-4 focused on identifying individuals at risk rather than diagnosing (and thus requiring functional impairment), consistent with our person-centered methodology. An individual's level of symptoms, and therefore their



risk, can fluctuate over time due to the interplay of personal factors (e.g., predispositions), contextual factors, and virtual contextual factors (e.g., precipitating, perpetuating, and protective factors; Stavropoulos, Motti-Stefanidi, & Griffiths, 2022). Consequently, functional impairment is not always necessary to assess level of risk. Instead, focusing on the subclinical risk thresholds of the GDT-4 as a self-report instrument is more appropriate as it reflects a targeted, prevention-oriented approach rather than a diagnostic claim. As prevention of mental disorders is both efficacious and cost-effective, this is an important goal (Mendelson & Eaton, 2018).

CONCLUSION

Thanks to Infanti et al. (2024), we had the opportunity to further elaborate on the theoretical aspects of the UAB definition and its GD risk predictive capacity. Although AI and ML techniques are still novel and promising analytical methods in the field of behavioral addictions, including by our team, we believe that the critical and inquisitive stance of Infanti et al. (2024) helps the field progress. On the contrary, the UAB's capacity to reveal information about the user, particularly their GD risk, has been well established over the past decade (e.g., Burleigh, Stavropoulos, Liew, Adams, & Griffiths, 2018; Stavropoulos, Gomez, Mueller, Yucel, & Griffiths, 2020), with ML methods adding a new layer of automation in decoding this information. In this context, we hope that our commentary response constructively promotes dialogue among scholars, while also equipping researchers with more knowledge regarding both the conceptual and methodological points raised. In this line, a corrigendum to the original article has been issued addressing the aforementioned concerns.

Funding sources: Vasileios Stavropoulos has received the Australian Research Council, Discovery Early Career Researcher Aw, Grant/Award Number: DE210101107.

Authors' contribution: Vasileios Stavropoulos wrote the first draft of this commentary response. Vasileios Stavropoulos, Mark D. Griffiths, Michelle Colder Carras, Daniel Zarate, Tyrone L. Burleigh, Bruno Schivinski, Leila Karimi, Angela Gorman-Alesi, Dylan Poulus, Taylor Brown, Rapson Gomez, Kaiden Hein, Rabindra Ratan, and Rachel Kowert contributed to the conceptual clarity around the UAB and around risk assessment in gaming disorder. Maria Prokofieva, Daniel Zarate, Vasileios Stavropoulos, and Nalin Arachchilage contributed to the methodological clarity part. All authors contributed to the writing and editing of the manuscript.

Conflicts of interest: MCC conducts scientific consulting on video games and has received grant support to conduct research on crisis prevention in online gaming communities. MDG has received research funding from *Norsk Tipping* (the gambling operator owned by the Norwegian

government). MDG has received funding for a number of research projects in the area of gambling education for young people, social responsibility in gambling and gambling treatment from *Gamble Aware* (formerly the *Responsibility in Gambling Trust*), a charitable body which funds its research program based on donations from the gambling industry. MDG undertakes consultancy for various gambling companies in the area of player protection and social responsibility in gambling.

Acknowledgment: The year 10 student Zakarij MCNARA contributed to initially identifying the points raised in the Infanti et al. (2024) commentary and shaping the form of the response politely and graciously during his work-experience week in the Associate Dean for HDR students office of the School of Health and Biomedical Sciences of RMIT University.

REFERENCES

- American Psychiatric Association (2022). Diagnostic and statistical manual of mental disorders (5th ed., text rev.). American Psychiatric Publishing. <https://doi.org/10.1176/appi.books.9780890425787>.
- Banks, J., & Bowman, N. D. (2021). Some assembly required: Player mental models of videogame avatars. *Frontiers in Psychology*, 12, 701965. <https://doi.org/10.3389/fpsyg.2021.701965>.
- Burleigh, T. L., Stavropoulos, V., Liew, L. W., Adams, B. L., & Griffiths, M. D. (2018). Depression, internet gaming disorder, and the moderating effect of the gamer-avatar relationship: An exploratory longitudinal study. *International Journal of Mental Health and Addiction*, 16, 102–124. <https://doi.org/10.1007/s11469-017-9806-3>.
- Casale, S., Musicò, A., Gualtieri, N., & Fioravanti, G. (2023). Developing an intense player-avatar relationship and feeling disconnected by the physical body: A pathway towards internet gaming disorder for people reporting empty feelings? *Current Psychology*, 42(24), 20748–20756. <https://doi.org/10.1007/s12144-022-03186-9>.
- Dar, A. W., & Farooq, S. U. (2024). Handling class overlap and imbalance using overlap driven under-sampling with balanced random forest in software defect prediction. *Innovations in Systems and Software Engineering*, 1–21. <https://doi.org/10.1007/s11334-024-00571-4>.
- Dawkins, R. (1982). *The extended phenotype: The long reach of the gene*. Oxford University Press.
- Eaton, W. W., Alexandre, P., Kessler, R. C., Martins, S. S., Mortensen, P. B., Rebok, G. W., ... Roth, K. (2012). The population dynamics of mental disorders. In W. W. Eaton (Ed.), *Public mental health* (pp. 125–150). Oxford University Press.
- Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, 73, 220–239. <https://doi.org/10.1016/j.eswa.2016.12.035>.
- Haraway, D. (2008). When species meet. In *The Routledge international handbook of more-than-human studies* (pp. 42–78). Routledge.

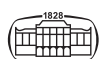


- He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)* (pp. 1322–1328). IEEE. <https://doi.org/10.1109/IJCNN.2008.4633969>.
- Infanti, A., Giardina, A., Razum, J., King, D. L., Baggio, S., Snodgrass, J. G., ... Billieux, J. (2024). User-avatar bond as diagnostic indicator for gaming disorder: A word on the side of caution. Commentary on: Deep learning(s) in gaming disorder through the user-avatar bond: A longitudinal study using machine learning (Stavropoulos et al., 2023). *Journal of Behavioral Addictions* (online advanced publication). <https://doi.org/10.1556/2006.2024.00032>.
- Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, M. (2021). Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. *IEEE Access*, 9, 39707–39716. <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9370099>.
- Khairy, M., Mahmoud, T. M., & Abd-El-Hafeez, T. (2024). The effect of rebalancing techniques on the classification performance in cyberbullying datasets. *Neural Computing and Applications*, 36(3), 1049–1065. <https://doi.org/10.1007/s00521-023-09084-w>.
- Loi, M. (2019). The digital phenotype: A philosophical and ethical exploration. *Philosophy & Technology*, 32(1), 155–171. <https://doi.org/10.1007/s13347-018-0319-1>.
- Mancini, T., Imperato, C., & Sibilla, F. (2019). Does avatar's character and emotional bond expose to gaming addiction? Two studies on virtual self-discrepancy, avatar identification and gaming addiction in massively multiplayer online role-playing game players. *Computers in Human Behavior*, 92, 297–305. <https://doi.org/10.1016/j.chb.2018.11.007>.
- Mendelson, T., & Eaton, W. W. (2018). Recent advances in the prevention of mental disorders. *Social Psychiatry and Psychiatric Epidemiology*, 53(4), 325–339. <https://doi.org/10.1007/s00127-018-1501-6>.
- Montag, C., & Rumpf, H. J. (2021). The potential of digital phenotyping and mobile sensing for psycho-diagnostics of internet use disorders. *Current Addiction Reports*, 8, 422–430. <https://doi.org/10.1007/s40429-021-00376-6>.
- Nowak, K. L., & Fox, J. (2018). Avatars and computer-mediated communication: A review of the definitions, uses, and effects of digital representations. *Review of Communication Research*, 6, 30–53. <https://doi.org/10.12840/issn.2255-4165.2018.06.01.015>.
- Ratan, R. A., & Dawson, M. (2016). When mii is me: A psychophysiological examination of avatar self-relevance. *Communication Research*, 43(8), 1065–1093. <https://doi.org/10.1177/0093650215570652>.
- Servidio, R., Griffiths, M. D., Boca, S., & Demetrovics, Z. (2023). The serial mediation effects of body image-coping strategies and avatar-identification in the relationship between self-concept clarity and gaming disorder: A pilot study. *Addictive Behaviors Reports*, 17, 100482. <https://doi.org/10.1016/j.abrep.2023.100482>.
- Stavropoulos, V., Gomez, R., Mueller, A., Yucel, M., & Griffiths, M. (2020). User-avatar bond profiles: How do they associate with disordered gaming? *Addictive Behaviors*, 103, 106245. <https://doi.org/10.1016/j.addbeh.2019.106245>.
- Stavropoulos, V., Motti-Stefanidi, F., & Griffiths, M. D. (2022). Risks and opportunities for youth in the digital era. *European Psychologist*, 27(2). <https://doi.org/10.1027/1016-9040/a000451>.
- Stavropoulos, V., Zarate, D., Prokofieva, M., Van De Berg, N., Karimi, L., Gorman Alesi, A., ... Griffiths, M. D. (2023). Deep learning(s) in gaming disorder through the user-avatar bond: A longitudinal study using machine learning. *Journal of Behavioral Addictions*, 12(4), 878–894. <https://doi.org/10.1556/2006.2023.00062>.
- Stavropoulos, V., Zarate, D., Prokofieva, M., Van De Berg, N., Karimi, L., Gorman Alesi, A., ... Griffiths, M. D. (2024). Corrigendum to: Deep learning(s) in gaming disorder through the user-avatar bond: A longitudinal study using machine learning. *Journal of Behavioral Addictions*, <https://doi.org/10.1556/2006.2024.40000>.
- Szolin, K., Kuss, D., Nuyens, F., & Griffiths, M. (2022). Gaming disorder: A systematic review exploring the user-avatar relationship in video games. *Computers in Human Behavior*, 128, 107124. <https://doi.org/10.1016/j.chb.2021.107124>.
- Tisseron, S. (2009). The teen and his avatars. *Adolescence*, 27(3), 591–600. <https://psycnet.apa.org/doi/10.3917/ado.069.0591>.
- World Health Organization (2022). *ICD-11: International classification of diseases* (11th revision). <https://icd.who.int/>.
- Zarate, D., Stavropoulos, V., Ball, M., De Sena Collier, G., & Jacobson, N. C. (2022). Exploring the digital footprint of depression: A PRISMA systematic literature review of the empirical evidence. *BMC Psychiatry*, 22(1), 421. <https://doi.org/10.1186/s12888-022-04013-y>.

Appendix

Results no SMOTE wave 1

Model	roc_auc	ppv	npv	f_meas	spec	sens	recall	accuracy
NBayes	0.513	0.802	0.500	0.855	0.045	0.988	0.988	0.796
R Forest	0.666	0.796	NaN	0.887	0.000	1.000	1.000	0.796
LASSO	0.692	0.619	0.617	0.887	0.000	1.000	1.000	0.796
Log reg	0.690	0.648	0.645	0.887	0.000	1.000	1.000	0.796
KNN	0.938	0.976	0.806	0.887	0.000	1.000	1.000	0.796
SVM	0.982	0.990	0.955	0.887	0.000	1.000	1.000	0.796
XGB	0.947	0.872	0.893	0.887	0.000	1.000	1.000	0.796



Results with SMOTE post splitting wave 1

Model	roc_auc	ppv	npv	f_meas	spec	sens	recall	accuracy
NBayes	0.573	0.806	0.300	0.859	0.802	0.689	0.919	0.759
R Forests	0.707	0.808	0.500	0.884	0.934	0.877	0.977	0.796
LASSO	0.665	0.796	NaN	0.887	0.896	0.811	1.000	0.796
Log reg	0.645	0.796	NaN	0.887	0.566	0.613	1.000	0.590
KNN	0.661	0.800	0.333	0.880	0.981	0.755	0.977	0.787
SVM	0.524	0.796	NaN	0.887	0.953	0.915	1.000	0.796
XG Boost	0.637	0.816	0.400	0.870	0.896	0.811	0.930	0.778

Results with ADASYN post splitting wave 1

Model	roc_auc	ppv	npv	f_meas	spec	sens	recall	accuracy
NBayes	0.583	0.812	0.231	0.723	0.802	0.689	0.651	0.602
R Forests	0.697	0.840	0.500	0.878	0.934	0.877	0.919	0.796
LASSO	0.633	0.836	0.255	0.694	0.585	0.642	0.593	0.583
Log reg	0.629	0.831	0.245	0.676	0.566	0.613	0.570	0.590
KNN	0.544	0.795	0.200	0.730	0.981	0.755	0.674	0.602
SVM	0.520	0.826	0.375	0.854	0.953	0.915	0.884	0.759
XG Boost	0.500	0.796	NaN	0.887	0.896	0.811	1.000	0.796

Results with ROSE post splitting wave 1

Model	roc_auc	ppv	npv	f_meas	spec	sens	recall	accuracy
NBayes	0.659	0.833	0.250	0.685	0.545	0.581	0.581	0.574
R Forests	0.631	0.864	0.286	0.703	0.636	0.593	0.593	0.602
LASSO	0.720	0.913	0.290	0.636	0.585	0.642	0.488	0.556
Log reg	0.720	0.913	0.290	0.636	0.818	0.488	0.488	0.590
KNN	0.586	0.863	0.263	0.642	0.682	0.512	0.512	0.546
SVM	0.629	0.896	0.283	0.642	0.773	0.500	0.500	0.556
XG Boost	0.500	0.796	NaN	0.887	0.000	1.000	1.000	0.796

Results with Tomek links post splitting wave 1

Model	roc_auc	ppv	npv	f_meas	spec	sens	recall	accuracy
NBayes	0.677	0.800	0.333	0.880	0.0455	0.977	0.977	0.787
R Forests	0.649	0.796	NaN	0.887	0.0000	1.000	1.000	0.796
LASSO	0.661	0.796	NaN	0.887	0.0000	1.000	1.000	0.796
Log reg	0.680	0.796	NaN	0.887	0.0000	1.000	1.000	0.590
KNN	0.648	0.796	NaN	0.887	0.0000	1.000	1.000	0.796
SVM	0.565	0.796	NaN	0.887	0.0000	1.000	1.000	0.796
XG Boost	0.672	0.796	NaN	0.887	0.0000	1.000	1.000	0.796

Results no SMOTE wave 2

Model	roc_auc	ppv	npv	f_meas	spec	sens	recall	accuracy
NBayes	0.612	0.792	0.167	0.840	0.083	0.894	0.894	0.729
R Forests	0.610	0.797	NaN	0.887	0.000	1.000	1.000	0.797
LASSO	0.710	0.797	NaN	0.887	0.000	1.000	1.000	0.797
Log reg	0.693	0.797	NaN	0.887	0.000	1.000	1.000	0.797
KNN	0.530	0.797	NaN	0.887	0.000	1.000	1.000	0.797
SVM	0.445	0.797	NaN	0.887	0.000	1.000	1.000	0.797
XG Boost	0.500	0.797	NaN	0.887	0.000	1.000	1.000	0.797



Results with SMOTE post splitting wave 2

Model	roc_auc	ppv	npv	f_meas	spec	sens	recall	accuracy
NBayes	0.624	0.829	0.278	0.773	0.417	0.723	0.723	0.661
R Forests	0.613	0.786	0.000	0.854	0.000	0.936	0.936	0.746
LASSO	0.500	0.797	N/A	0.887	0.000	1.000	1.000	0.797
Log reg	0.704	0.797	N/A	0.887	0.000	1.000	1.000	0.590
KNN	0.587	0.797	N/A	0.887	0.000	1.000	1.000	0.797
SVM	0.560	0.797	N/A	0.887	0.000	1.000	1.000	0.797
XG Boost	0.500	0.797	N/A	0.887	0.000	1.000	1.000	0.797

Results with ADASYN post splitting wave 2

Model	roc_auc	ppv	npv	f_meas	spec	sens	recall	accuracy
NBayes	0.658	0.848	0.269	0.700	0.583	0.596	0.596	0.593
R Forests	0.631	0.808	0.286	0.848	0.167	0.894	0.894	0.746
LASSO	0.500	0.797	NaN	0.887	0.000	1.000	1.000	0.797
Log reg	0.695	0.914	0.375	0.78	0.750	0.681	0.681	0.695
KNN	0.491	0.800	0.211	0.736	0.333	0.681	0.681	0.610
SVM	0.486	0.792	0.182	0.800	0.167	0.809	0.809	0.678
XG Boost	0.500	0.797	NaN	0.887	1.000	0.797	0.500	0.797

Results with ROSE post splitting wave 2

Model	roc_auc	ppv	npv	f_meas	spec	sens	recall	accuracy
NBayes	0.613	0.885	0.273	0.630	0.750	0.489	0.489	0.542
R Forests	0.649	0.848	0.269	0.700	0.583	0.596	0.596	0.593
LASSO	0.681	0.829	0.250	0.707	0.500	0.617	0.617	0.593
Log reg	0.681	0.829	0.250	0.707	0.500	0.617	0.617	0.590
KNN	0.550	0.812	0.222	0.658	0.500	0.553	0.553	0.542
SVM	0.670	0.857	0.292	0.732	0.583	0.638	0.638	0.627
XG Boost	0.500	0.797	NaN	0.887	0.000	1.000	1.000	0.797

Results with Tomek links post splitting wave 2

Model	roc_auc	ppv	npv	f_meas	spec	sens	recall	accuracy
NBayes	0.534	0.804	0.200	0.874	0.0833	0.915	0.915	0.780
R Forests	0.656	0.797	NaN	0.887	0.0000	1.000	1.000	0.797
LASSO	0.716	0.797	NaN	0.887	0.0000	1.000	1.000	0.797
Log reg	0.670	0.797	NaN	0.887	0.0000	1.000	1.000	0.651
KNN	0.571	0.797	NaN	0.887	0.0000	1.000	1.000	0.797
SVM	0.301	0.797	NaN	0.887	0.0000	1.000	1.000	0.797
XG Boost	0.500	0.797	NaN	0.887	0.0000	1.000	1.000	0.797

Open Access statement. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium for non-commercial purposes, provided the original author and source are credited, a link to the CC License is provided, and changes – if any – are indicated.

