



LJMU Research Online

Agius, S, Cassar, V, Magri, C, Khan, W and Topham, L

Position paper: advocating for a structured methodology in developing data-driven predictive models for healthcare – evidence from a large-scale national study

<https://researchonline.ljmu.ac.uk/id/eprint/26943/>

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

**Agius, S, Cassar, V, Magri, C, Khan, W ORCID logoORCID:
<https://orcid.org/0000-0002-7511-3873> and Topham, L ORCID logoORCID:
<https://orcid.org/0000-0002-6689-7944> (2025) Position paper: advocating for
a structured methodology in developing data-driven predictive models for**

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

<http://researchonline.ljmu.ac.uk/>



Position paper: advocating for a structured methodology in developing data-driven predictive models for healthcare – evidence from a large-scale national study

Stephen Agius¹ · Vincent Cassar¹ · Caroline Magri² · Wasiq Khan³ · Luke Topham³

Received: 3 July 2025 / Accepted: 31 July 2025
© The Author(s) 2025

Abstract

Background Despite the growing adoption of predictive models in healthcare, the development process is often inconsistent and lacks methodological rigour. Many models are created ad hoc, without transparent handling of missing data, proper validation, or alignment with clinical workflows. These shortcomings have undermined trust, reproducibility, and generalisability, especially in high-stakes environments like emergency care.

Objectives This position paper aims to advocate for the adoption of structured, transparent, and reproducible methodologies in the development of predictive models for healthcare. Drawing on a large-scale national study of emergency department (ED) visits in Malta, the paper demonstrates that methodological discipline, guided by data science principles, clinical expertise and an understanding of human decision-making behaviour leads to safer, more trustworthy, and clinically relevant models.

Methods Using over 32 million data points from 650,000 ED visits across six years, the study employed a structured modelling pipeline that integrated clinical and administrative data sources. The methodology included Cognitive Task Analysis (CTA) to map triage decision-making, rigorous feature engineering based on clinical workflows, handling of missing data through informed strategies, and robust model validation using XGBoost with stratified cross-validation and calibration analysis. Importantly, domain experts were involved throughout the development lifecycle to ensure clinical relevance and interpretability.

Results The structured methodology enabled the development of predictive models that reflected the real-world complexity of ED triage, achieved strong performance, and gained clinician acceptance. The models aligned with staged clinical decision-making and were interpretable, trustworthy, and feasible to scale across healthcare environments. Through transparent documentation, robust calibration, and post-deployment monitoring protocols, the models demonstrated readiness for clinical integration.

Conclusions The study confirms that structured, domain-informed methodologies are not only feasible at scale but essential for the responsible deployment of predictive models in healthcare. This approach ensures safety, fosters trust, promotes reproducibility and increases the likelihood that the model is used and adopted in real clinical settings. The authors call on researchers, developers, and regulators to establish such methodologies as the standard for AI and data-driven approaches in healthcare, particularly in high-stakes applications where poor model performance can lead to clinical harm.

Keywords Digital health technologies · Predictive models · Clinical decision · Healthcare automation · Data-driven decision support · Health information technology · Digital transformation in healthcare · Healthcare quality improvement

1 Introduction

Predictive models are increasingly recognised as powerful tools for enhancing clinical decision-making across diverse healthcare settings [19]. Their ability to uncover hidden patterns, support early interventions, and optimise resource allocation is driving widespread adoption [5, 28]. In triage environments, where clinicians must make rapid decisions under uncertainty, data-driven models can augment clinical judgement and reduce variability [4, 13]. Recent evidence underscores how big-data and machine learning (ML) are reshaping patient triage and operational forecasting in emergency care, with such technologies improving accuracy and reducing under-triage of high-risk patients, thereby enabling more timely clinical responses [1].

This position paper draws on the findings of a previously conducted large-scale national study to argue that the lack of structured methodologies in predictive model development remains a critical barrier to safe and effective adoption in healthcare. Inconsistent practices in data handling, feature selection, model validation, and transparency continue to undermine trust and generalisability [8, 18, 20, 26]. In clinical settings, where predictive models directly influence patient care, such deficiencies can pose serious safety risks, erode clinical trust, and potentially lead to harm [29].

This position paper contends that structured, reproducible methodologies must become the standard in healthcare predictive modelling. We support this position with insights from a large-scale national study of emergency department (ED) visits conducted across Malta's entire healthcare system. The study analysed more than 32 million data points derived from over 650,000 patient visits to the national ED over a six-year period [2]. This unprecedented scale enabled us to examine the complex, real-world decision-making processes in triage and to empirically demonstrate how structured, reproducible methodologies can enhance the safety, transparency, and clinical relevance of predictive models in such critical environments.

2 Problem statement

Despite growing interest in predictive models, many are developed in ad hoc ways [18, 30]. Common deficiencies include:

- Poor documentation of data sources and preprocessing [11].
- Limited attention to handling missing data [12, 15].
- Inadequate validation and calibration [9, 27].
- Lack of transparency and explainability [10, 11].

- Limited attention to behavioural factors shaping clinical decision-making [23].

These issues hinder reproducibility and create models that may perform well in narrow settings but fail when deployed more broadly. In healthcare, this can undermine patient safety, erode clinician trust, and hinder the effective adoption of data-driven and AI-enabled systems.

3 Position statement

We take the position that any predictive model used in healthcare, especially in time-critical settings such as triage, must be developed through a structured, transparent, and reproducible methodology, grounded in both data science best practices and clinical relevance.

4 Supporting arguments

The argument for adopting structured, reproducible methodologies in healthcare predictive modelling is grounded both in foundational principles of data science and in practical insights drawn from our large-scale study of ED triage. In this section, we articulate key elements that are essential to building clinically relevant and trustworthy models, from data quality and integration to feature engineering, model validation, and lifecycle management. Each of these elements was informed not only by the technical demands of machine learning, but also by our Cognitive Task Analysis (CTA) of triage decision-making, which underscored the need to align predictive models with the cognitive and operational realities of clinical practice [3]. The following supporting arguments demonstrate why generic, unstructured approaches are insufficient in healthcare, and why structured methodologies must become the standard.

4.1 Data quality is foundational

The value of a predictive model depends fundamentally on the quality and relevance of its input data [21]. Our study used retrospective patient-level data from Malta's national Health Information System, encompassing over 653,000 ED visits between 2017 and 2022, supplemented by laboratory data.

A structured methodology for data identification ensures that models reflect the complexity and realities of clinical decision-making, capturing both initial triage variables and later diagnostic indicators.

4.2 Unified data structuring enables holistic modelling

Raw healthcare data are often fragmented, originating from disparate systems such as electronic health records, laboratory information systems, and administrative databases [22]. This fragmentation can lead to inconsistencies, redundancies, and missing links in the patient journey. To address this, our methodology emphasised rigorous data integration, harmonising variables across multiple platforms using a common patient identifier. This ensured that all data points whether clinical, diagnostic, or administrative were correctly attributed to the same individual across time and setting.

Such unification is essential for building robust predictive models capable of learning meaningful patterns across the entire continuum of care, from entry at triage to final discharge. Without comprehensive integration, models risk being trained on incomplete or biased representations of patient trajectories, which can undermine performance, validity, and generalisability, particularly in high-stakes environments like emergency care [14, 16].

4.3 Addressing missing data transparently

Missing data is a common and often unavoidable challenge in complex data environments such as healthcare, where not all diagnostic tests or clinical inputs are performed or recorded uniformly [25]. If not appropriately addressed, missing data can introduce bias, reduce statistical power, and compromise the validity and generalisability of predictive models. Structured and transparent strategies for handling missing data are therefore essential to maintain model reliability and clinical relevance.

Predictive models can approach missing data in several ways, including deletion of incomplete records, statistical imputation, model-based handling, or categorical encoding. In our study, we adopted a dual strategy tailored to both the data structure and the clinical context:

- a) Numerical variables (e.g., lab values) were imputed using a default value of zero, allowing the model to process the data consistently without removing cases. This choice was based on model compatibility and the fact that missing values in our context often reflected “not tested” rather than an error.
- b) Categorical variables (e.g., test result labels) were encoded using a distinct ‘MISSING’ category. This allowed the model to learn patterns associated with the absence of information, which can itself be informative for example, when the non-performance of a test may correlate with triage urgency or clinical suspicion.

This strategy enables machine learning algorithms to treat missingness as a potentially informative feature rather than as noise, thereby preserving the integrity of the model’s predictive patterns. Importantly, transparent and consistent handling of missing data is essential to maintain trust and accountability in any data-driven application.

4.4 Feature engineering must reflect clinical reality

Feature engineering bridges data science and clinical practice. In our study, variables were selected to mirror real-world triage decision-making, including presenting complaints, entry method, age, seasonality, and blood biomarkers.

We adopted a two-stage modelling approach aligned with clinical workflows:

- Stage 1: initial triage decisions based on limited data.
- Stage 2: enhanced predictions once richer diagnostics became available.

Such alignment ensures that models support rather than disrupt clinical processes, as it mimics the real-life process of clinical diagnostics, where initial assessments are refined as more information becomes available.

A key principle guiding our methodology is the systematic integration of clinical domain knowledge at every stage of model development. Throughout the project, expert input from ED clinicians, triage nurses, and hospital operations staff was actively sought and incorporated. Domain experts contributed to the initial selection of clinically meaningful variables, advised on the relevance and interpretation of laboratory markers, and validated data transformations to ensure alignment with real-world clinical workflows. In feature engineering, clinicians provided insight into the temporal dynamics of triage decisions, informed the grouping of complaint categories, and contextualised blood biomarker thresholds. Importantly, during model evaluation, feature importance plots and calibration curves were reviewed in joint clinical, data science workshops to assess whether the model’s decision logic aligned with clinical reasoning and patient safety expectations. This continuous dialogue between data and behaviour scientists and healthcare professionals not only enhanced the model’s interpretability but also fostered trust and readiness for future clinical integration. Our experience underscores that domain knowledge should not be confined to feature selection but embedded throughout the model development lifecycle to ensure that predictive models augment rather than disrupt clinical judgement.

4.5 Rigorous model selection and validation are essential

Model selection must be driven by both statistical performance and clinical interpretability. We chose XGBoost for its ability to handle structured data, missing values, and class imbalance [7].

Our training process included:

- An 80:20 train-test split - divided the dataset into 80% for training the model and 20% for evaluating its performance on unseen data
- SMOTE for balancing rare classes - applied Synthetic Minority Over-sampling Technique to generate synthetic samples for under-represented classes in the data.
- Hyperparameter tuning - Optimised model parameters systematically to achieve the best predictive performance.
- 5-fold stratified cross-validation - assessed model performance by training and testing on five data folds while preserving class distribution in each fold.

Evaluation went beyond accuracy, using calibration plots to assess probability estimation, a critical factor in risk-based clinical decision-making [27].

Explainability was enhanced through feature importance plots, enabling clinicians to understand the drivers of model predictions. Without such transparency, models cannot earn trust or regulatory approval.

4.6 Model monitoring and retraining

Predictive models in healthcare operate in dynamic environments where patient populations, clinical practices, and operational workflows evolve over time. To maintain clinical relevance and safeguard patient safety, it is essential to implement robust processes for ongoing model monitoring and retraining. In our proposed methodology, we advocate for continuous performance tracking post-deployment using statistical process control techniques, calibration drift monitoring, and periodic evaluation of model outputs against clinical outcomes. Alerts should be generated if performance metrics such as AUC-ROC, calibration, or recall degrade beyond predefined thresholds. Furthermore, retraining protocols must be established in advance, specifying triggers for initiating retraining (e.g., substantial population shift or emergence of new clinical guidelines), versioning of models, and documentation requirements. Importantly, retraining should once again involve close collaboration with domain experts to ensure that any updated model remains aligned with clinical expectations and regulatory standards. This lifecycle perspective recognises that

in healthcare, model governance must extend well beyond initial development and deployment.

5 Counterarguments and rebuttals

- a) Counterargument - Structured methodologies are too slow for innovation in healthcare AI.

Rebuttal - In healthcare, speed must never come at the expense of clinical safety, or patient trust. While structured methodologies may initially require more deliberate effort, they provide a foundation for models that are not only more accurate, but also more transparent, explainable, and aligned with clinical realities. Our study demonstrated that by systematically integrating domain knowledge and adhering to rigorous validation, it is possible to accelerate the safe deployment of predictive models. In contrast, generic models often encounter resistance from clinicians, fail regulatory scrutiny, and require costly rework. In the long term, structured processes reduce risk, enhance reproducibility, and foster innovation that is sustainable and scalable. In a domain where lives are at stake, rigour is not an obstacle to innovation, it is its enabler.

- b) Counterargument - Off-the-shelf models or autoML tools are sufficient.

Rebuttal - While automation can assist in accelerating certain aspects of model development, healthcare predictive models require careful alignment with clinical workflows and transparent handling of complex data challenges, factors that generic automated tools often fail to address [6, 17, 20, 24, 26]. Our research combined large-scale ED data analysis with CTA to explicitly map how triage decisions are made under real-world conditions of time pressure, information uncertainty, and clinical judgement. This process highlighted the importance of capturing not only structured clinical data, but also the implicit reasoning and contextual factors that shape decision-making. Feature engineering, model validation, and interpretability were therefore closely informed by domain knowledge elicited through CTA. Generic tools are ill-equipped to handle this level of contextual and cognitive nuance. As such, structured, domain-specific methodologies grounded in an understanding of actual clinical practice such as those developed in this study, are essential to building predictive models that clinicians can trust, adopt, and use to enhance patient care.

6 Implications of adopting structured methodologies

Adopting structured methodologies in healthcare predictive modelling carries significant implications for clinical practice, policy, and research. For clinical practice, structured approaches will improve clinician trust in predictive models by enhancing transparency and reliability. They will enable more accurate and actionable decision support, directly contributing to better patient outcomes, and will help reduce variability and the potential for harm when models are deployed in real-world settings.

For policymakers, structured methodologies should become a regulatory requirement. Regulators should mandate transparent documentation of model development processes, explicit handling and reporting of data quality issues and missing data strategies, and robust validation and calibration procedures to ensure models are safe and effective when implemented in clinical environments.

For researchers, structured methodologies provide a clear roadmap for advancing the field of healthcare AI. They support the development of reproducible models, promote the open sharing of models and results, and contribute to the responsible evolution of healthcare predictive modelling as a scientific discipline.

7 Conclusion

In healthcare, predictive models hold the potential to transform patient care, enhancing decision-making, anticipating clinical deterioration, and optimising resource allocation. However, their impact depends not only on technical performance but on how they are developed, validated, and integrated into real-world clinical workflows. As demonstrated in our large-scale national study of emergency department triage, methodological rigour is not a luxury, it is a prerequisite for safety, transparency, and clinical acceptance.

We contend that structured methodologies rooted in data science best practices and informed by clinical expertise are essential for building trustworthy predictive models. From data integration and feature engineering to transparent handling of missing data, validation, and continuous monitoring, every stage of model development must be conducted deliberately and systematically. This is particularly vital in high-stakes environments like emergency care, where flawed models can compromise patient safety and clinician confidence.

Our experience in analysing over 32 million data points across 650,000 ED visits shows that structured approaches are not only feasible at scale but also yield models that are clinically relevant, interpretable, and better aligned with

the decision-making realities of frontline healthcare professionals. Moreover, by embedding cognitive and contextual insights through methods such as CTA and by combining behavioural science with technology, we demonstrate that structured methodologies can bridge the gap between data-driven insights and clinical intuition.

We therefore call on the broader healthcare AI community, researchers, developers, regulators, and journal editors to adopt and enforce structured methodologies as a standard, not an option. This includes mandating transparent documentation, rigorous validation, and meaningful clinical engagement throughout the model lifecycle. Only then can predictive models truly deliver on their promise: to support clinicians, safeguard patients, and contribute to the responsible and equitable advancement of data-driven healthcare.

Author contributions SA conceptualised and designed the study. VC, CM and WK contributed to data collection. LT assisted in the interpretation of the results and drafted parts of the manuscript. All authors reviewed and approved the final version of the manuscript.

Funding Open Access funding provided by the University of Malta. This research is part of a PhD that was funded by the Tertiary Education Scholarship Scheme managed by the Ministry of Education, Malta.

Data availability The data generated and analysed during the study (interview transcripts and observation notes) are not publicly available due to privacy and ethical restrictions.

Code availability No custom code or software was developed for this study. Qualitative data analysis was performed using NVivo (Release 1.7.1), a commercially available software package.

Declarations

Ethics approval This study was conducted in accordance with the Declaration of Helsinki and received approval from the University of Malta Research Ethics Committee (Ref: FEMA-2023-00285) and the Data Protection Office at Mater Dei Hospital (Ref: 34/2023).

Consent to participate Written informed consent was obtained from all triage nurses who participated in interviews, as well as from nurses and patients involved in observational sessions.

Consent for publication All participants provided consent for the anonymised publication of study findings. No individual participant is identifiable in the article or any associated materials.

Conflict of interest No conflict of interest has been declared by the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless

indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Agius S, Cassar V, Bezzina F, Topham L. (2025). Leveraging digital technologies to enhance patient safety. *Health and Technology*, 0123456789. <https://doi.org/10.1007/s12553-025-01001-6>
2. Agius S, Cassar V, Magri C, et al. Predicting Emergency Severity Index (ESI) level, hospital admission, and admitting ward in an emergency department using data-driven machine learning. *BMC Med Inform Decis Mak*. 2025;25:281. <https://doi.org/10.1186/s12911-025-02941-9>.
3. Agius S, Magri C, Cassar V. A cognitive task analysis for developing a clinical decision support system for emergency triage. *J Emerg Nurs*. 2025. <https://doi.org/10.1016/j.jen.2025.05.013>.
4. Barak-Corren Y, Israelit SH, Reis BY. (2017). Progressive prediction of hospitalisation in the emergency department: uncovering hidden patterns to improve patient flow. *Emergency Medicine Journal*, 34(5), 308 LP – 314. <https://doi.org/10.1136/emered-2014-203819>
5. Brnabic A, Hess LM. Systematic literature review of machine learning methods used in the analysis of real-world data for patient-provider decision making. *BMC Med Inf Decis Mak*. 2021;21(1). <https://doi.org/10.1186/s12911-021-01403-2>.
6. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA*. 2017;318(6):517–8. <https://doi.org/10.1001/jama.2017.7797>.
7. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proc ACM SIGKDD Int Conf Knowl Discovery Data Min*. 2016;13–17–Augu:785–94. <https://doi.org/10.1145/2939672.2939785>.
8. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. 2020;21(1):6. <https://doi.org/10.1186/s12864-019-6413-7>.
9. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *Lancet*. 2019;393(10181):1577–9. [https://doi.org/10.1016/S0140-6736\(19\)30037-6](https://doi.org/10.1016/S0140-6736(19)30037-6).
10. Doshi-Velez F, Kim B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. *ML*, 1–13. <http://arxiv.org/abs/1702.08608>
11. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health*. 2021;3(11):e745–50. [https://doi.org/10.1016/S2589-7500\(21\)00208-9](https://doi.org/10.1016/S2589-7500(21)00208-9).
12. Haneuse S, Daniels M. A general framework for considering selection bias in EHR-Based studies: what data are observed and why EGEMs (Generating Evid Methods Improve Patient Outcomes). 2016;4(1):16. <https://doi.org/10.13063/2327-9214.1203>.
13. Harrison CJ, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction to natural Language processing. *BMC Med Res Methodol*. 2021;21(1):1–18. <https://doi.org/10.1186/s12874-021-01347-1>.
14. Hripesak G, Albers DJ. Next-generation phenotyping of electronic health records. *J Am Med Inform Assoc*. 2013;20(1):117–21. <https://doi.org/10.1136/amiajnl-2012-001145>.
15. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Med Res Methodol*. 2017;17(1):162. <https://doi.org/10.1186/s12874-017-0442-1>.
16. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. 2012;13(6):395–405. <https://doi.org/10.1038/nrg3208>.
17. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17(1):195. <https://doi.org/10.1186/s12916-019-1426-2>.
18. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *BMJ*. 2020;370:m3164. <https://doi.org/10.1136/bmj.m3164>.
19. Niu S, Ma J, Yin Q, Wang Z, Bai L, Yang X. Modelling patient longitudinal data for clinical decision support: a case study on emerging AI healthcare technologies. *Inform Syst Front*. 2024. <https://doi.org/10.1007/s10796-024-10513-x>.
20. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347–58. <https://doi.org/10.1056/nejmr1814259>.
21. Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, Moons KGM, Collins G, Van Smeden M. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368(March):1–12. <https://doi.org/10.1136/bmj.m441>.
22. Salmund SW, Echevarria M. Healthcare transformation and changing roles for nursing. *Orthop Nurs*. 2017;36(1):12–25. <https://doi.org/10.1097/NOR.0000000000000308>.
23. Schwartz J, George M, Rossetti S, Dykes P, Minshall S, Lucas E, Cato K. Factors influencing clinician trust in predictive clinical decision support systems for in-hospital deterioration: qualitative descriptive study. *JMIR Hum Factors*. 2022;9(2):e33960. <https://doi.org/10.2196/33960>.
24. Sendak MP, Ratliff W, Sarro D, Alderton E, Futoma J, Gao M, Nichols M, Revoir M, Yashar F, Miller C, Kester K, Sandhu S, Corey K, Brajer N, Tan C, Lin A, Brown T, Engelbosch S, Anstrom K, O'Brien C. Real-World integration of a sepsis deep learning technology into routine clinical care: implementation study. *JMIR Med Inf*. 2020;8(7):e15182. <https://doi.org/10.2196/15182>.
25. Steyerberg EW. (2019). Clinical prediction models. In *Statistics in the health sciences*.
26. Topol E. Deep medicine: how artificial intelligence can make healthcare human again. Basic Books; 2019.
27. Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW, Bossuyt P, Collins GS, MacAskill P, Moons KGM, Vickers AJ. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17(1):1–7. <https://doi.org/10.1186/s12916-019-1466-7>.
28. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, Cumbers S, Jonas A, McAllister KSL, Myles P, Granger D, Birse M, Branson R, Moons KGM, Collins GS, Ioannidis JPA, Holmes C, Hemingway H. (2020). Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *The BMJ*, 368. <https://doi.org/10.1136/bmj.l6927>.
29. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, Jung K, Heller K, Kale D, Saeed M, Ossorio PN, Thadaneysrani S, Goldenberg A. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med*. 2019;25(9):1337–40. <https://doi.org/10.1038/s41591-019-0548-6>.
30. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, Albu E, Arshi B, Bellou V, Bonten MMJ, Dahly DL, Damen JA, Debray TPA, de Jong VMT, De Vos M, Dhiman P, Ensor J, Gao S, Haller MC, van Smeden M. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*. 2020;369:m1328. <https://doi.org/10.1136/bmj.m1328>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Stephen Agius¹  · Vincent Cassar¹ · Caroline Magri² · Wasiq Khan³ · Luke Topham³

✉ Stephen Agius
stephen.agius@um.edu.mt

Vincent Cassar
vincent.cassar@um.edu.mt

Caroline Magri
caroline.j.magri@um.edu.mt

Wasiq Khan
W.Khan@ljmu.ac.uk

Luke Topham
L.K.Topham@ljmu.ac.uk

¹ Department of Business and Enterprise Management, Faculty of Economics, Management & Accountancy, University of Malta, Msida, Malta

² Mater Dei Hospital, Msida, Malta

³ Liverpool John Moores University, Liverpool, England