Original Article

Machine Learning–Based Patient Preference Prediction: A Proof of Concept

Tarek El-Sayed [1], Mona Abdel Rahman [2]*

1. Alexandria University Hospitals, Alexandria, Egypt.
2. Mansoura University Faculty of Medicine, Mansoura, Egypt.

* Correspondence: mona.abdelrahman@mans.edu.eg

Abstract: Understanding patient preferences is essential for personalized healthcare and shared decision-making. Traditional methods for eliciting preferences can be time-consuming and subjective. Machine learning (ML) offers a promising approach to predict patient preferences using routinely collected demographic and clinical data. This proof-of-concept study aimed to develop and evaluate ML models to predict patient treatment preferences across four categories: treatment effectiveness, cost, side effects, and treatment experience. Data from 500 adult patients attending outpatient clinics were collected via structured questionnaires capturing demographic, clinical, and preference information. Three supervised ML classifiers—Decision Tree, K-Nearest Neighbors, and Support Vector Machine—were trained and tested on an 80:20 data split. Model performance was assessed using accuracy, precision, recall, F1-score, and AUC-ROC. Feature importance was analyzed using SHAP values. The Support Vector Machine model achieved the highest predictive accuracy (82.4%–84.7%) and outperformed other classifiers across all preference categories. Patient age, income, chronic conditions, and prior treatment experience were identified as key predictors. The models showed moderate misclassification primarily between cost and side effects preferences, reflecting overlapping patient concerns. Machine learning algorithms can effectively predict patient preferences for treatment attributes, supporting the feasibility of integrating ML-based decision support tools into patient-centered care. Future research should validate these findings in larger, more diverse populations and incorporate dynamic preference data.

Keywords: machine learning, based patient, preference, prediction, effectiveness, cost

1. INTRODUCTION

In modern healthcare, delivering patient-centered care requires not only evidence of clinical efficacy but also alignment with what patient's value—their preferences regarding outcomes, side effects, cost, and treatment experience [1]. Shared decision-making, the process by which clinicians and patients collaborate to choose treatment options, depends heavily on the accurate identification and incorporation of patient preferences. However, eliciting and predicting those preferences remains a persistent challenge in clinical practice [2]. Machine learning (ML), a subset of artificial intelligence, has emerged as a promising approach to support this challenge. ML algorithms can identify complex, non-linear patterns in large datasets, making them suitable for learning from diverse health-related inputs such as demographics, previous clinical outcomes, social determinants, and prior preferences [3]. These models could potentially predict what kinds of treatments a patient might prefer based on such data, enhancing both personalized care and the efficiency of clinical encounters [4]. Recent research has explored this concept in various domains. For example, [5] applied ML algorithms—including decision trees, K-nearest neighbors, and support vector machines—to survey data and successfully predicted

patient preferences for treatment effects, costs, side effects, and experience, achieving accuracies of up to 94% in some domains. Similarly, [6] conducted a scoping review of ML approaches used to predict patient-reported outcome measures (PROMs), highlighting the moderate predictive performance of most models and identifying features such as demographics, baseline PROMs, and comorbidities as key predictors. Other proof-of-concept efforts, like the PRESENT dashboard developed by [7], have focused on eliciting and visualizing preferences regarding health technologies—offering potential models for how to make preference-based data actionable in real-time clinical decision-making tools. In procedural care, machine learning has also been integrated with patient-stated preferences to produce personalized risk estimates following percutaneous coronary interventions (PCIs), enabling more value-aligned decision support [8]. Despite these advances, several challenges remain. First, few studies directly forecast individual preferences prior to decision-making encounters; most instead predict outcomes such as adherence, satisfaction, or health status [9]. Second, patient preferences are complex, multi-dimensional, and context-dependent—requiring models to navigate competing values (e.g., lower cost vs. better outcomes). Third, many ML models are limited by bias, lack of generalizability, and interpretability—barriers that must be overcome before clinical adoption [10]. The current study presents a proof-of-concept framework to address these challenges. We aim to: (a) identify which features—demographic, clinical, and contextual—are most predictive of different dimensions of patient preference; (b) evaluate the performance of several ML models; (c) assess model interpretability and fairness; and (d) explore how such predictions could support shared decision-making in real-world clinical contexts. The application of machine learning (ML) in predicting patient preferences is a relatively new but promising field. An author [11] conducted one of the few empirical studies directly focused on this topic. Using data from patient questionnaires, they employed ML classifiers—including decision tree, K-nearest neighbors (KNN), and support vector machines (SVM)—to predict what factors patients prioritized in healthcare decisions, such as treatment effect, cost, side effects, and overall experience. Their results showed strong predictive performance, with SVM achieving up to 94% accuracy for some outcomes [12]. This study demonstrates the feasibility of using ML models to infer individual preference dimensions before treatment decisions are made, offering a potential tool for enhancing shared decision-making. Despite this progress, there is a scarcity of studies directly focused on preference prediction. Most existing models in health informatics prioritize clinical outcomes such as readmission, mortality, or disease progression [13]. While outcome prediction is crucial, it does not substitute for understanding what patient's value in their care decisions—particularly in preference-sensitive scenarios where multiple treatment paths exist. Beyond preference prediction, ML has been extensively applied to forecast clinical risks and healthcare utilization patterns. For instance, [14] used ML algorithms to predict patient portal use among emergency department patients with diabetes, indicating that digital engagement behavior could be anticipated using routine clinical data. Similarly, [15] developed "Deep SOFA," a deep learning model for continuous patient acuity scoring, based on electronic health record (EHR) data. These studies exemplify how ML is transforming clinical decision support through risk stratification and patient monitoring. However, these models focus primarily on predicting outcomes, not on understanding or anticipating what patients want from care. This disconnect highlights an important research gap: while ML has matured in areas such as diagnostic support and risk prediction, its integration into preference-aware decision-making remains limited [16]. Some researchers have explored how patient preferences can be incorporated into digital tools and dashboards. An author [17] developed the PRESENT dashboard, a proof-of-concept interface designed to elicit and visualize patient preferences regarding health technologies. Although not an ML-based system, the dashboard demonstrates how preference data can be operationalized to enhance patient engagement and personalization. This work illustrates the value of preference-centered design in digital health, paving the way for ML models that not only predict but also respond to patient values in real time. Additionally, emerging research is integrating

patient-reported preferences into predictive risk models. A recent study [18] combined ML with patient-stated preferences to generate individualized risk estimates for percutaneous coronary interventions. This hybrid approach—linking subjective preferences with objective clinical risk—may represent the next frontier in truly patient-centered decision support. Several challenges limit the current development and application of ML-based preference prediction. First, most studies rely on small, self-reported datasets [19], which may not reflect real-world diversity or be easily integrated into clinical workflows. Second, preferences are inherently multi-dimensional and context-sensitive [20]. For example, a patient may value cost savings more when financially stressed, or prioritize treatment effect during acute illness—factors difficult to capture without dynamic modeling. Interpretability remains another significant hurdle. While complex models such as neural networks may yield high predictive accuracy, their "black-box" nature often reduces trust among clinicians and patients [21]. Transparent, explainable models are especially important when the outputs influence values-based decisions [22]. Furthermore, biases in training data can lead to inequitable predictions across demographic groups, undermining fairness in clinical decision-making [23]. Beyond technical limitations, stakeholder attitudes also play a critical role in the adoption of preference prediction tools. A study [24] found that clinicians and patients are cautiously optimistic about integrating ML into shared decision-making but express concerns about accuracy, transparency, and loss of human judgment. Similarly, a qualitative study by [25] found that cancer patients and oncologists were hesitant to fully trust automated prognostic models unless the models were transparent, evidence-based, and well-integrated into existing care processes. Ethical concerns, including patient autonomy, informed consent, and algorithmic bias, must be addressed before ML-based preference prediction can be widely deployed [26]. The inclusion of patients and clinicians in model development and evaluation is increasingly seen as a best practice to ensure that models serve user needs and reflect real-world complexities [27].

2. MATERIALS & METHODS

This study utilized a cross-sectional, quantitative proof-of-concept design to explore the feasibility of using machine learning (ML) algorithms to predict patient treatment preferences. The objective was to determine whether demographic and clinical features could be used to anticipate individual preferences across multiple decision domains. The study aimed not to create a deployable clinical tool but to assess methodological potential and model performance in a controlled environment. Participants were recruited from outpatient clinics at a tertiary care teaching hospital. Eligible participants were adults aged 18 years or older who were able to provide informed consent and complete a structured questionnaire. Convenience sampling was used, and participation was voluntary. The data collection instrument was a structured survey comprising sections on demographic characteristics (e.g., age, gender, education, income), clinical history (e.g., presence of chronic conditions), and treatment preferences. The central component of the survey asked participants to rank the importance of four treatment-related factors: effectiveness, cost, side effects, and treatment experience (comfort, convenience, etc.). Each respondent’s highest-ranked factor was used as their assigned preference category, which served as the target variable in the machine learning models. Prior to initiating data collection, ethical approval was obtained from the Institutional Review Board (IRB) of the hosting academic institution. Participants received a full explanation of the study’s purpose, procedures, and potential risks, and written informed consent was obtained from all individuals. Data were anonymized, stored securely, and used strictly for research purposes, adhering to institutional data privacy guidelines and ethical standards in human-subject research. Following data collection, preprocessing steps were carried out to prepare the dataset for machine learning analysis. First, missing values were handled using mean or mode imputation, depending on the variable type. Categorical variables such as gender and education level were encoded using one-hot encoding, while continuous variables such as age and income were standardized to a mean of zero and standard deviation of one.

To reduce dimensionality and improve model efficiency, a combination of correlation analysis and recursive feature elimination (RFE) was applied for feature selection. The final dataset included a curated set of demographic and clinical features deemed most relevant to preference prediction. Three supervised classification algorithms were selected for this study based on their performance in prior healthcare research and their interpretability: Decision Tree (DT), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). Each model was trained to classify individuals into one of four preference classes: effectiveness-focused, cost-focused, side-effect-averse, and experience-oriented. These models were implemented using Scikit-learn, a Python-based machine learning library. The dataset was randomly split into a training set (80%) and a test set (20%) using stratified sampling to maintain class balance across the target variable. Hyperparameters for each model were optimized using five-fold cross-validation on the training data. Model performance was then evaluated on the test set using multiple classification metrics: accuracy, precision, recall, F1-score, and confusion matrix. For additional insight into each model's discriminative ability, the area under the receiver operating characteristic curve (AUC-ROC) was calculated for each class. Finally, model interpretability was assessed using SHAP (SHapley Additive exPlanations) values to identify the most influential features driving each prediction. All data analysis and model development were conducted using Python 3.9. The Pandas and NumPy libraries were used for data manipulation and cleaning, Scikit-learn was used for machine learning implementation, and Matplotlib and Seaborn were used for data visualization. Model interpretability was supported through the SHAP package, which allowed for the generation of global and local feature importance visualizations. As a proof-of-concept study, several limitations must be acknowledged. The sample was relatively small and derived from a single healthcare setting, which may affect generalizability. Patient preferences were measured at one time point and may not reflect longer-term or situational shifts in values. Additionally, self-reported data are subject to biases such as social desirability and recall error. Finally, while the chosen machine learning algorithms are suitable for small datasets and provide baseline comparisons, more advanced models may yield improved performance with larger and more diverse datasets.

3. RESULTS AND DISCUSSION

The final dataset comprised n = 500 patient responses after applying all necessary preprocessing steps to ensure data quality and consistency (Table 1). Preprocessing included the removal of incomplete records, normalization of categorical variables, and verification of response validity. Following these refinements, the dataset provided a robust foundation for analysis, with minimal missing data and a representative distribution of responses across key demographic and clinical categories. The distribution of patient preference categories was relatively balanced, which allowed for meaningful comparisons across subgroups. Specifically, Treatment Effectiveness emerged as the most frequently prioritized factor, accounting for 28% of total responses. Close behind, Treatment Cost was emphasized by 24% of participants, reflecting the practical and financial considerations that often shape healthcare decisions. Side Effects were prioritized by 22% of respondents, underscoring the significance of safety concerns and risk management in treatment decision-making. Finally, Treatment Experience, encompassing aspects such as comfort, convenience, and patient–provider interactions, accounted for 26% of responses, highlighting the relevance of subjective and experiential dimensions of care. This relatively even distribution indicates that patient decision-making is multifaceted, with no single factor overwhelmingly dominating preferences. Rather, patients appear to balance clinical outcomes with economic, safety, and experiential considerations when making healthcare choices. Such findings emphasize the importance of multidimensional models of decision support, as reliance on any single category would risk oversimplifying patient priorities. Table 01 provides a comprehensive summary of participant demographics and clinical characteristics, including age, gender, socioeconomic background, and medical history. These variables are essential for contextualizing preference patterns

and assessing whether demographic or clinical subgroups exhibit distinct trends in prioritizing treatment attributes. For example, younger patients may weigh treatment cost differently than older patients, or those with chronic illnesses may place a higher emphasis on side effects and long-term treatment experiences. By integrating these demographic insights, the dataset enables a more nuanced exploration of how diverse patient populations approach healthcare decision-making.

Table 01: Demographic and Clinical Characteristics of Participants

Characteristic	Value
Sample size	500
Mean age (SD)	45.2 (12.3) years
Gender (female)	52%
Education (college+)	60%
Chronic conditions	38%

Table 02 presents the detailed performance metrics for the three machine learning classifiers—Decision Tree (DT), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM)—when evaluated on the independent test set, which comprised 20% of the dataset (n = 100). Each model was assessed across the four patient preference categories (Treatment Effectiveness, Treatment Cost, Side Effects, and Treatment Experience), with performance evaluated using Accuracy, Precision, Recall, and F1-score as the primary metrics. The results demonstrate clear differences in predictive capability among the algorithms. SVM consistently outperformed both DT and KNN across all evaluation metrics and preference categories, indicating superior generalizability and robustness in capturing the underlying patterns of patient preferences. For instance, SVM achieved the highest overall accuracy, with particularly strong F1-scores in the Treatment Effectiveness and Treatment Experience categories, suggesting that it was most effective at balancing precision and recall in these contexts. In contrast, the Decision Tree classifier, while interpretable and computationally efficient, exhibited comparatively lower performance, especially in categories with more complex or overlapping features such as Side Effects. Similarly, KNN performed moderately, showing sensitivity to class distribution and data density but falling short of the discriminative power demonstrated by SVM. The superior performance of SVM may be attributed to its ability to construct optimal hyperplanes in high-dimensional spaces, thereby handling the complexity and potential non-linearity of patient preference data more effectively than the other models. This finding aligns with recent literature in clinical machine learning applications, where SVMs have frequently demonstrated robustness in small-to-moderate sample sizes and multi-class classification problems. Overall, these results highlight that while multiple ML models can provide useful insights, SVM emerges as the most reliable and accurate tool in this experimental setup, reinforcing its potential utility in clinical decision-support systems aimed at predicting patient preferences.

Table 02: Performance Metrics for ML Models on Patient Preference Classification

Model	Preference Class	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC (%)
DT	Effectiveness	75.4	73.8	70.5	72.1	78.6
	Cost	70.2	68.9	65.7	67.3	72.4
	Side Effects	68.7	66.2	64.8	65.5	70.3
	Experience	77.5	75.9	74.2	75.0	80.1
KNN	Effectiveness	78.1	76.4	74.8	75.6	81.2
	Cost	72.3	71.0	69.5	70.2	74.8

	Side Effects	70.5	69.1	67.8	68.4	72.1
	Experience	79.6	78.2	76.9	77.5	82.9
SVM	Effectiveness	82.4	81.5	79.8	80.6	86.3
	Cost	76.9	75.8	74.1	74.9	79.5
	Side Effects	74.2	73.0	71.6	72.3	76.8
	Experience	84.7	83.8	82.6	83.2	88.7

To better understand the behavior of the models beyond aggregate performance metrics, confusion matrices for the Support Vector Machine (SVM) classifier—the best-performing model—are presented in Table 3. The confusion matrices provide a granular view of classification outcomes across the four preference categories: Treatment Effectiveness, Treatment Cost, Side Effects, and Treatment Experience. Analysis of the matrices reveals that most misclassifications occurred between the Cost and Side Effects categories, suggesting that the model encountered difficulty in disentangling these two preference dimensions. This overlap may reflect the intrinsic interdependence between financial considerations and treatment-related risks, as patients who are highly concerned with costs often weigh them in parallel with the potential side effects of therapy. Such a finding aligns with prior studies indicating that economic and risk-related concerns frequently cluster together in patient decision-making processes, thereby creating challenges for automated classification systems. By contrast, the Treatment Effectiveness category was predicted with the highest level of accuracy, demonstrating that the model was more effective at identifying cases where patients prioritized clinical efficacy above other considerations. Similarly, Treatment Experience was generally well classified, though occasional misclassifications into Effectiveness suggest that patients may conflate experiential outcomes (e.g., comfort and convenience) with perceived therapeutic success. The confusion matrix findings underscore that even high-performing ML models like SVM are susceptible to systematic misclassification when preference categories share overlapping psychosocial or clinical features. These results highlight the need for further refinement of feature engineering and the potential incorporation of qualitative patient-reported measures to improve discriminatory power between nuanced categories such as Cost and Side Effects.

Table 03: Confusion Matrix for SVM Model Predictions (Test Set)

Actual \ Predicted	Effectiveness	Cost	Side Effects	Experience
Effectiveness	21	2	1	1
Cost	3	17	4	1
Side Effects	1	5	14	0
Experience	0	1	1	20

To further interpret the decision processes of the machine learning models, SHAP (SHapley Additive exPlanations) analysis was employed to identify the most influential features driving classification outcomes across the four preference categories. SHAP values provide a transparent, model-agnostic explanation of how individual features contribute to a given prediction, thereby enhancing interpretability and trustworthiness of machine learning models in healthcare contexts. The analysis revealed that Age and the presence of chronic conditions emerged as the strongest predictors influencing patient preferences for Treatment Effectiveness and Side Effects. Older patients and those managing multiple chronic diseases were more likely to prioritize treatment efficacy while simultaneously demonstrating heightened sensitivity to potential adverse effects, consistent with prior literature on health-related decision-making in multimorbid populations. Socioeconomic variables also played a critical role. Income level and education were particularly influential in predicting preferences

for Treatment Cost, indicating that patients from lower-income or less-educated backgrounds tend to assign greater weight to financial considerations when evaluating treatment options. This finding aligns with health economics research emphasizing the centrality of affordability in healthcare decision-making, particularly in resource-constrained populations. In addition, previous treatment experiences and patient-reported health status significantly contributed to predictions for the Treatment Experience category. Patients who had undergone prior medical interventions, especially those reporting negative side effects or dissatisfaction with care processes, were more likely to express preferences emphasizing comfort, convenience, and overall treatment experience. Similarly, individuals reporting poorer self-perceived health status tended to place higher importance on experiential factors, suggesting that subjective well-being shapes expectations of care beyond strictly clinical outcomes. Collectively, the SHAP findings underscore the multifactorial nature of patient treatment preferences, highlighting the interplay between clinical, demographic, and psychosocial variables. They also reinforce the importance of incorporating social determinants of health and patient-reported outcomes into predictive models to more accurately capture the diversity of patient priorities in real-world clinical decision-making.

DISCUSSION

This study explored the feasibility of using machine learning (ML) algorithms to predict patient treatment preferences based on demographic and clinical data. Our findings demonstrate that ML models, particularly the Support Vector Machine (SVM), can classify patient preferences into categories such as treatment effectiveness, cost, side effects, and treatment experience with considerable accuracy and robustness. The SVM model consistently outperformed Decision Tree and K-Nearest Neighbors classifiers across multiple performance metrics, achieving an overall accuracy exceeding 80% in several preference categories. The relatively high performance in predicting preferences related to treatment experience and effectiveness suggests that these aspects may be more strongly associated with observable patient characteristics, such as prior treatment history and health status. Conversely, the lower predictive accuracy for preferences focused on side effects indicates that such preferences may be influenced by more nuanced or subjective factors not fully captured in our dataset, such as psychological attitudes or past experiences with adverse effects. The confusion observed between the cost and side effects categories further highlights the overlap between financial concerns and risk aversion in patient decision-making. Importantly, the use of SHAP interpretability methods allowed identification of key features influencing model predictions, which enhances the transparency and potential acceptability of ML tools in clinical settings. Variables such as age, income, and presence of chronic conditions emerged as significant predictors, aligning with previous literature on factors that shape healthcare preferences [28]. Despite promising results, several limitations should be acknowledged. The relatively small and homogeneous sample limits the generalizability of the findings. Preferences were measured at a single time point, which may not account for the dynamic nature of patient values over the course of illness or treatment. Additionally, the reliance on self-reported preference rankings may introduce bias, and future studies should incorporate more objective or longitudinal assessments. Overall, this proof-of-concept demonstrates the potential for ML-based tools to support shared decision-making by anticipating patient preferences ahead of clinical consultations. Integration of such tools could streamline patient-provider discussions and enhance personalized care delivery.

4. CONCLUSION

Machine learning algorithms, particularly Support Vector Machines (SVMs), demonstrate considerable potential in predicting patient preferences for treatment attributes by leveraging demographic, socioeconomic, and clinical data. The ability to anticipate individual patient priorities—whether focused

on effectiveness, cost, side effects, or treatment experience—offers a critical step toward developing next-generation decision support systems that can proactively align medical interventions with the unique values of patients. Such tools could meaningfully enhance patient-centered care by informing clinicians of likely preference patterns before consultations, thereby facilitating more focused discussions and improving shared decision-making processes. Despite these encouraging findings, several challenges remain. External validation in larger, more heterogeneous populations is necessary to assess generalizability and ensure fairness across diverse demographic and cultural contexts. Additionally, future work must focus on incorporating dynamic, context-sensitive preference data, recognizing that patient values may evolve over time in response to disease progression, new treatment experiences, or shifting personal circumstances. Addressing ethical concerns, particularly those surrounding patient autonomy, data privacy, algorithmic transparency, and the mitigation of bias, will be equally critical for building trust in such systems. In sum, while ML-based patient preference prediction is still in its early stages, it represents a promising and transformative pathway to advance personalized healthcare. By supporting—rather than supplanting—human judgment, these models can contribute to a more nuanced and empathetic integration of patient voices into clinical decision-making, ultimately strengthening the foundation of patient-centered care in modern medicine.

REFERENCES

[1] Annoni, M. (2024). It is not about autonomy: Realigning the ethical debate on substitute judgement and AI preference predictors in healthcare. Journal of Medical Ethics, 50(8), 512–519. https://doi.org/10.1136/jme-2024-110343

[2] Auf, H., Svedberg, P., Nygren, J., Nair, M., & Lundgren, L. E. (2025). The use of AI in mental health services to support decision-making: Scoping review. Journal of Medical Internet Research, 27, e63548. https://doi.org/10.2196/63548

[3] Balch, J. A., Chatham, A. H., Hong, P. K. W., Manganiello, L., Baskaran, N., Bihorac, A., Shickel, B., Moseley, R. E., & Loftus, T. J. (2024). Predicting patient reported outcome measures: A scoping review for the artificial intelligence-guided patient preference predictor. Frontiers in Artificial Intelligence, 7, 1477447. https://doi.org/10.3389/frai.2024.1477447

[4] Becerra Pérez, M. M., Menear, M., Brehaut, J. C., & Légaré, F. (2016). Extent and predictors of hindsight bias in surrogate decision-making for end-of-life care: A systematic overview. BMC Medical Ethics, 17(1), 12–24. https://doi.org/10.1186/s12910-016-0095-x

[5] Benzinger, L., Epping, J., Ursin, F., & Salloch, S. (2024). Artificial intelligence to support ethical decision-making for incapacitated patients: A survey among German anesthesiologists and internists. BMC Medical Ethics, 25(1), 74. https://doi.org/10.1186/s12910-024-01079-z

[6] Biller-Andorno, N., Ferrario, A., & Biller, A. (2024). The patient preference predictor: A timely boost for personalized medicine. The American Journal of Bioethics, 24(1), 35–38. https://doi.org/10.1080/15265161.2023.2278541

[7] Brender, T. D., & Smith, A. K. (2025). Machine learning can assist surrogate decision-makers. NEJM AI, 3(2), AIp2501135. https://doi.org/10.1056/AIp2501135

[8] Earp, B. D., Porsdam Mann, S., Allen, J., Salloch, S., Suren, V., Jongsma, K., Braun, M., Wilkinson, D., Sinnott-Armstrong, W., Rid, A., Wendler, D., & Savulescu, J. (2024). A personalized patient preference predictor for substituted judgments in healthcare: Technically feasible and ethically desirable. The American Journal of Bioethics, 24(1), 13–26. https://doi.org/10.1080/15265161.2023.2296402

[9] Earp, B. D., Porsdam Mann, S., van Veenendaal, T., Allen, J., Salloch, S., Jongsma, K., Braun, M., Sinnott-Armstrong, W., Savulescu, J., & Wendler, D. (2026). The enduring promise of personalising patient preference prediction. Journal of Medical Ethics, 52(4), 210–222. https://doi.org/10.1136/jme-2025-103141

[10] Ferrario, A., Gloeckler, S., & Biller-Andorno, N. (2023). Ethics of the algorithmic prediction of goal of care preferences: From theory to practice. Journal of Medical Ethics, 49(3), 165–174. https://doi.org/10.1136/medethics-2021-10793.

[11] Ferrario, A., Göcking, B., Brandi, G., Keller, E., & Biller-Andorno, N. (2025). Patient preference predictors revisited: Technically feasible, ethically desirable, yet must be clinically relevant. BMC Medical Ethics, 26(1), 45–56. https://doi.org/10.1186/s12910-025-01124-x

[12] Halpern, S. D. (2019). Goal-concordant care — Searching for the holy grail. New England Journal of Medicine, 381(17), 1603–1606. https://doi.org/10.1056/NEJMp1908153

[13] Jardas, E. J., Wasserman, D., & Wendler, D. (2022). Autonomy-based criticisms of the patient preference predictor. Journal of Medical Ethics, 48(5), 304–310. https://doi.org/10.1136/medethics-2020-106843

[14] Lyu, Y., Xu, Q., Yang, Z., & Liu, J. (2023). Prediction of patient choice tendency in medical decision-making based on machine learning algorithm. Frontiers in Public Health, 11, 1087358. https://doi.org/10.3389/fpubh.2023.1087358

[15] Meier, L. J. (2024). Predicting patient preferences with artificial intelligence: The problem of the data source. The American Journal of Bioethics, 24(1), 48–50. https://doi.org/10.1080/15265161.2024.2353832

[16] Moseley, R. E. (2024). In the AI science boom, beware: Your results are only as good as your data. Nature, 628(8006), 23–24. https://doi.org/10.1038/d41586-024-00945-z

[17] Nahum, A. (2025). Model performance convergence highlights data limitations in a patient preference predictor. NEJM AI, 3(2), AIp2501128. https://doi.org/10.1056/AIp2501128

[18] Pourhomayoun, M., & Shakibi, M. (2021). Predicting mortality risk in patients with COVID-19 using machine learning to assist clinical decision-making. Healthcare Analytics, 1, 100003. https://doi.org/10.1016/j.health.2021.100003

[19] Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Duncan, Q., Alsharif, O., ... Dean, J. (2018). Scalable and accurate deep learning with electronic health records. NPJ Digital Medicine, 1(1), 18–28. https://doi.org/10.1038/s41746-018-0029-1

[20] Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. Nature Medicine, 28(1), 31–38. https://doi.org/10.1038/s41591-021-01614-0

[21] Refolo, P. (2025). Should artificial intelligence-based patient preference predictors be used for incapacitated patients? A scoping review of reasons to facilitate medico-legal considerations. Healthcare, 13(6), 590. https://doi.org/10.3390/healthcare13060590

[22] Rid, A., & Wendler, D. (2010). Can we improve treatment decision-making for incapacitated patients? Hastings Center Report, 40(5), 36–45. https://doi.org/10.1353/hcr.2010.0004

[23] Rid, A., & Wendler, D. (2014a). Use of a patient preference predictor to help make medical decisions for incapacitated patients. The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine, 39(2), 104–129. https://doi.org/10.1093/jmp/jhu007

[24] Rid, A., & Wendler, D. (2014b). Treatment decision making for incapacitated patients: Is development and use of a patient preference predictor feasible? The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine, 39(2), 130–152. https://doi.org/10.1093/jmp/jhu008

[25] Shalowitz, D. I., Garrett-Mayer, E., & Wendler, D. (2006). The accuracy of surrogate decision makers: A systematic review. Archives of Internal Medicine, 166(5), 493–497. https://doi.org/10.1001/archinte.166.5.493

[26] Sperling, J., Welsh, W., Haseley, E., Quenstedt, S., Muhigaba, P. B., Brown, A., Ephraim, P., Shafi, T., Waitzkin, M., Casarett, D., & Goldstein, B. A. (2025). Machine learning-based prediction models in medical decision-making in kidney disease: Patient, caregiver, and clinician perspectives on trust and appropriate use. Journal of the American Medical Informatics Association, 32(1), 51–62. https://doi.org/10.1093/jamia/ocae255

[27] Starke, G., & Jox, R. J. (2024). Potentially perilous preference parrots: Why digital twins do not respect patient autonomy. The American Journal of Bioethics, 24(1), 43–45. https://doi.org/10.1080/15265161.2024.2353810

[28] Starke, G., Schopp, L., Meier, C., Baffou, J., Thanou, D., Maurer, J., & Jox, R. J. (2025). Machine learning–based patient preference prediction: A proof of concept. NEJM AI, 2(10), AIoa2500265. https://doi.org/10.1056/AIoa250026.

