A Glimpse Into The Future: The Ethical Implications of Predicting Patient Mortality

Ava Klissouras

10/16/24

Predicting the future is often perceived as a superpower, an unattainable omniscience able to be possessed only within a fictional realm. However, thanks to the power of machine learning algorithms, this fantasy ability may, in fact, exist within our human grasp—at least, in the case of predicting mortality. The journal article "Development of a prognostic model for mortality in COVID-19 infection using machine learning" by Adam L. Booth, Elizabeth Abels, and Peter McCaffrey, develops a machine learning algorithm that predicts mortality from COVID-19 up to 48 hours prior to patient expiration. However, this paper raises significant ethical concerns regarding the moral soundness of predicting death. In America's bleak, overstretched, and underfunded medical landscape—which can be unexpectedly puncuated by medical miracles and unpredicted recoveries—does attempting to gaze into a patient's future—to declare their fate before time has run its course—have the capacity to overshadow serendipitous medical suprises and do more harm than good?

"Development of a prognostic model for mortality in COVID-19 infection using machine learning" (Booth et al.) presents a retrospective study, which evaluates laboratory data and mortality from 398 COVID-positive patients at the University of Texas Medical Branch. The authors retrospectively searched the hospital's Laboratory Information System, and located

values for twenty-six serum chemistry and blood gas laboratory parameters in the patients. The data included all patients admitted in the hospital with positive COVID test results— the rationale for this was that there would be fewer missing laboratory test results if only admitted patients (not just all patients with positive tests), were included, thereby lessening the need for later imputation of the data. The authors discarded metrics for which less than 25% of patients had measured values within 14 days following a positive COVID-19 test. Additionally, they excluded results captured within 48 hours of death in order to ensure the sufficiently early identification of patients who are likely to pass.

Once the authors had obtained their data, they performed multivariate feature imputation using Scikit-Learn's IterativeImputer method. This technique models each feature with missing values as a function of the other features, and computes an estimate for the missing value using that model. This method is still in its experimental phase– developers note that it is not yet fully stable/finalized, and its default parameters may change without notifying users. The authors employed Bayesian Ridge Regression as their model, and included the expired versus non-expired variable as one of the inputs in this model. This inclusion was motivated by their desire to preserve correlations between each of the imputed variables, and patient mortality.

After imputation, the authors partitioned the data into training and testing sets, and performed a logistic regression to determine which features were most important in predicting mortality (based on regression coefficients). They then chose the five features with the highest weights to be used for the machine learning component of the analysis, with the intention of preserving parsimony. These features were C-reactive protein (CRP; indicates inflammation in the body), blood urea nitrogen (BUN), serum calcium, serum albumin, and lactic acid. The authors then used these features to train a Support Vector Machine (SVM) to predict patient mortality, theorizing that the features could have nonlinear interactions, and consequentially employing a radial basis kernel. However, the authors acknowledg that the increased model performance obtained by utilizing a nonlinear kernel comes at the expense of

2

model interpretability. In order to mitigate this, the authors implemented Shapley additive explanations (SHAP), a technique that increases interpretability by capturing the marginal contribution of each feature to the model's ultimate output. The authors preemptively mitigated the impact of imputed values in SHAP analysis by dropping sparse measurements, as described above.

The authors use a variety of metrics to evaluate the effectiveness of their models. The initial logistic regression that the authors used to select the five most "important" features achieved 80% sensitivity and 77% specificity on the testing data set. Moreover, the SVM trained using these features achieved 91% sensitivity and 91% specificity on the testing data. The positive and negative predicted values (PPV and NPV) are 62.5% and 98.4%, respectively (where patient expiration is classified as a positive outcome). The fact that NPV exceeds PPV indicates that a higher proportion of the non-mortality predictions are correct, compared to the mortality predictions. This makes sense, considering mortality is the minority class in this analysis, and for a rare outcome, NPV typically exceeds PPV (as negative (non-mortality) predictions are more likely to be true).

Additionally, the authors calculated Shapley values for each feature to determine the relative influence they exert on the model predictions. CRP, lactic acid, and calcium have the highest and lowest Shapley values, and thus exert the strongest influence on model output. The Shapley values also communicate some more complex results about the influence exerted by features. For instance, BUN plays a larger role in predicting mortality when CRP, lactic acid, and calcium do not take on significantly abnormal values in patients. Additionally, nonlinear relationships exist between albumin and lactic acid, and CRP and calcium. CRP influences model outcome only when calcium is elevated, and albumin is influential when lactic acid is elevated. Increasingly abnormal values of CRP exert greater influence on model predictions, except in cases with high BUN, where CRP's effect decreases. Altogether, these results communicate interactions between the features, which make it clear that the prediction of mortality is more complex than a simple additive relationship.

3

The results of this paper generate a variety of ethical concerns, perhaps the most pressing being the potential impact on the quality of healthcare individuals receive when their mortality can be predicted. The authors of this paper note that prediction of mortality can facilitate better allocation of critical care supplies and staff. However, this seemingly positive statement begs the question, *what is meant by "better"?* One lens through which to view this statement is the allocation of resources away from individuals who are predicted to pass away, and instead towards patients who are expected to survive. After all, if these patients are to pass anyway, why not allocate as many resources as possible towards preventing more deaths? Based on principles of deontology, this use of the authors' machine learning model would be inherently immoral. According to deontologists, the justifiability (or lack thereof) of an action is completely determined by whether it aligns with the two formulations of the categorical imperative: (1) whether it can be willed into universal law without logical contradiction, and (2) whether it treats humanity as a means, rather than a mere end. The use of the authors' algorithm to disparately provide care based on patients' predicted mortality status violates the second formulation of the categorical imperative, as the individuals who involuntarily relinquish the resources that would be allocated to them as a result of their predicted mortality are treated as means to an end. Moreover, if these individuals are neglected by doctors, who may choose to care for patients who are not predicted to pass away, they will be robbed of the ability to experience a medical miracle and survive.

Another interpretation of the "better" allocation of critical care supplies and staff is the allocation of resources towards patients who are predicted to pass away, motivated by the fact that they are in more dire need of care. However, this also violates the second formulation of the categorical imperative, as it treats patients who are expected to survive as means to an end. They are involuntarily deprived of resources that could have been used for their care, which are instead used to preserve the lives of patients who are predicted to pass.

Violation of just one of the two formulations of the categorical imperative is sufficient to render an act immoral, and in both of the above scenarios, the second formulation of the

categorical imperative is violated. Thus, the application of the algorithm developed in this paper is not morally sound, based on principles of deontology.

It is worthwhile to note that the algorithm presented in this paper could potentially have benefits that outweigh its moral costs: namely, if it allows for an optimized allocation of resources that minimizes patients' overall mortality due to COVID. If this were the case, then based on principles of Utilitarianism, which declares that the morality of an act is determined by whether it maximizes pleasure and minimizes pain, the implementation of this algorithm would be morally sound. In this case, pleasure could be interpreted as patient survival/recovery, in addition to easing the burden of medical professionals, and pain could be interpreted as patient mortality, and patients not receiving the care that they need.

In summary, the mortality-predicting SVM developed by Booth et al. has the potential to be a useful tool with potential to improve the efficiency of medical care, allowing professionals to optimize the allocation of resources towards those who would benefit the most from them. However, it also raises the moral dilemma of choosing exactly who these individuals are. Is it ever ethical to intentionally deny one patient care over another? Is it ethical to seek the knowledge that allows us to do so? Perhaps using machine learning to gain a glimpse into patients' futures is a double-edged sword: one that has the capacity to help some, at the expense of harming others.

References

Booth, A. L., Abels, E., & McCaffrey, P. (2021). Development of a prognostic model for mortality in COVID-19 infection using machine learning. *Modern Pathology*, 34(3), 522–531. https://doi.org/10.1038/s41379-020-00700-x

Sandel, M. J. (2010). *Justice: What's the right thing to do?* Farrar, Straus and Giroux.

6.4. imputation of missing values. Scikit Learn. (n.d.). https://scikit-learn.org/1.5/modules/impute.html