Evaluating Machine Learning Techniques for Early PTSD prognosis after Trauma using synthetic data.

Author:

Thomas J Wise

Student Number:

6664202

Programme:

Methodology and Statistics for the Behavioural, Biomedical and Social Sciences

Supervisors:

Dr. Mirjam van Zuiden

Prof. dr. Rens van de Schoot

Proposed Journal of Publication:

Multivariate Behavioural Research (MBR)

Ethics Approval

Status Accepted: #20-0039; Date: 09/10/2020

Status Accepted: #20-0140; Date: 27/10/2020

Introduction

One of the potential psychopathological consequences of exposure to traumatic events (accidents, mass violence, war or natural disasters) is Post-Traumatic Stress Disorder (PTSD, Breslau, Davis, Andreski, & Peterson, 1991). PTSD is characterized through an array of physical and psychological symptoms which culminate in a significant impairment to daily life (Association & others, 2013; Augsburger & Galatzer-Levy, 2020). Condition prognosis, is not considered adequately described through the use binary diagnostic criteria. Instead though prognosis trajectories, not reliant upon diagnostic thresholds, but rather the severity and course of symptoms over time (Bonanno, 2004; Bonanno & Diminich, 2013). Current research supports the presence of four dominant trajectories which define PTSD prognosis: resilient, recovered, chronic and delayed (Bonanno, 2004; Galatzer-Levy, Huang, & Bonanno, 2018; Schoot et al., 2018).

Understanding these trajectories, and their predictive factors is important, given the potential effectiveness of early, targeted PTSD interventions (Freedman, Eitan, & Weiniger, 2020; Kearns, Ressler, Zatzick, & Rothbaum, 2012; Short, Morabito, & Gilmore, 2020). However, to effectively provide these targeted interventions after a traumatic event, a sufficiently validated, trajectory focused, screening instrument is required. Current research has presented several screening instruments (SPAN (Meltzer-Brody, Churchill, & Davidson, 1999; Zlotnick, Davidson, Shea, & Pearlstein, 1996); IES-R (Marmer & Weiss, 1997); and TSQ (Foa, Riggs, Dancu, & Rothbaum, 1993)) which can predict with adequate accuracy (>0.8 Area under the Curve), PTSD diagnosis after 6 months (SPAN=0.83, IES-R=0.83, TSQ=0.82; Mouthaan, Sijbrandij, Reitsma, Gersons, & Olff, 2014). These however, predict whether a participant will be above the clinical threshold for PTSD, rather than their specific trajectory. Although these examples are not exhaustive, there are currently no validated trajectory focused screening instruments currently in clinical use.

To develop such an instrument, it is essential to build a statistical model; as without it can be challenging to understand the role and function of individual predictors (Taylor, Ford, & Ford, 2010). In the present study, a statistical model is required to 1) appropriately classify individuals to their specific trajectory and 2) predict this membership using predictors which are available directly after a traumatic event. To develop this model, firstly the application of a bayesian latent growth mixture model (outlined in: Zuiden et al., n.d.) is required to estimate the trajectories of participants in the existing data. This estimation uses a participants change in Clinician Administered PTSD Scale (CAPS) scores (Blake et al., 1995) across multiple time points to estimate their trajectory. In the present study, this estimation is used to provide the true trajectories for which the predicted membership are compared with allowing the development of Performance Metrics.

Secondly, predictors which will be available directly after a traumatic event (see Selection of Variables of Interest) are used as part of Machine Learning (ML) techniques to provide predictions for these trajectory memberships. The ML technique, and appropriate model derived, will than be used to develop an instrument using the most significant predictors from those selected available.

Supervised ML techniques are used development of this statistical model, given their capacity to identify complex underlying patters within data. In particular, this research will evaluate 9 ML techniques (see Table 1). These have been demonstrated as effective in PTSD prognosis research (Galatzer-Levy, Karstoft, Statnikov, & Shalev, 2014; Ramos-Lima, Waikamp, Antonelli-Salgado, Passos, & Freitas, 2020) and across the more general prognosis fields of psychiatric (Passos et al., 2016; Webb et al., 2020) and physical (Cruz & Wishart, 2006; Kononenko, 2001; Kourou, Exarchos, Exarchos, Karamouzis, & Fotiadis, 2015; Rajkomar, Dean, & Kohane, 2019) medicine. A broad range of literature across prognosis research was evaluated to identify suitable ML techniques, to reduce the potential influence of publication bias arising through the isolated examination of only PTSD prognosis research.

Therefore the current research will address the question: Which Machine Learning technique is most appropriate for predicting early PTSD prognosis early after Trauma, given the inherently limited sample size within the field?

To fully address this question, several methodological issues are accounted for. Firstly, to uphold the statistical integrity of the existing data, models will be developed and tested upon synthetic simulated data. This synthetic data will use the population parameters provided by this existing data, and be generated using one of three potential simulation methods (discussed in: Data Simulation). Through the application of ML techniques on this synthetic data, rather than the existing data, this will reduce the violations to statistical integrity which result from multiple exploratory testing (Ranganathan, Pramesh, & Buyse, 2016). The use of synthetic data, generated as part of a simulation, also provides benefits to understanding the influence of sample size when comparing ML techniques. As typically in PTSD prognosis research, data is presented with inherently limited sample sizes. For example, the existing data comprises of 852 cases (Mouthaan et al., 2014), whereas other research studies have examined trajectories with small-to-moderate sample sizes (min=15, max=1765, mean=343; Smid, Mooren, Mast, Gersons, & Kleber, 2009) and the largest individual study currently proposed, being of 5,000 participants (McLean et al., 2020). The restrictions resulting from these sample sizes, such as reduced reliability and increased cross validation errors (Varoquaux, 2018), can be overcome through the use of synthetic data. Where sample sizes can be varied, to determine the influence of the sample size on the accuracy of the model produced, to determine the optimal sample size to evaluate a specific model or the most appropriate model for a specific sample size. This specific use of synthetic data is widely supported, where prognostic and statistical models can be developed and tested during conditions when real data is unable to be used solely (Ambler, Brady, & Royston, 2002; Burton, Altman, Royston, & Holder, 2006; Morris, White, & Crowther, 2019).

Methodological & Analytical Approach

In the following subsections, corresponding to those in Figure 1, the methodological and analytical approach is outlined.

Data, Overview of the Existing Database

The TraumaTips data set (Mouthaan et al., 2014) provides the population parameters for data simulation. This data is used as it accurately represent the wider PTSD population of recently traumatized individuals regarding subsequent PTSD development (Shalev et al., 2019).

Collected between September 2005 and March 2009, this data was collected to investigate the prevalence and course of PTSD after traumatic injury, and compare the validity of different prognosis screening instrument of PTSD. This data consists of 852 patients who were presenting at two level-1 trauma centers in Amsterdam (NL). Patients were excluded from the existing study if their injuries were due to self-harm, they presented with an organic brain condition, current psychotic symptoms or associated disorder, bipolar disorder or depression with psychotic features, moderate to severe traumatic brain injury or held permanent residency outside of the Netherlands. Participants were assessed at 5 time points (T1; 23 days (mean), T2; 1 month, T3; 3 months, T4; 6 months and T5; 12 months), after admission to the study. Of relevance to the current study, at T1 participants completed a range of self-report scales related to PTSD and potential risk factors, in addition to PTSD screening instruments (including SPAN (Meltzer-Brody et al., 1999; Zlotnick et al., 1996), IES-R (Marmer & Weiss, 1997), TSQ (Foa et al., 1993)). At T3-T5, participants completed a clinician administered semi-structured interview, CAPS, based upon the DSM-IV criteria (Blake et al., 1995).

Prior to data simulation, several data cleaning methods were applied including: A) Reclassification of Variables, B) Removal of Retrospective Questionnaires, and C) Time Based-Grouping. These methods, replicated those applied in previous research (Zuiden et al., n.d.), and make the data suitable for calculating a participants trajectory membership. Membership is calculated using a bayesian latent growth mixture model formula, applied post-simulation to the CAPS scores to provide the true membership classification for the synthetic data.

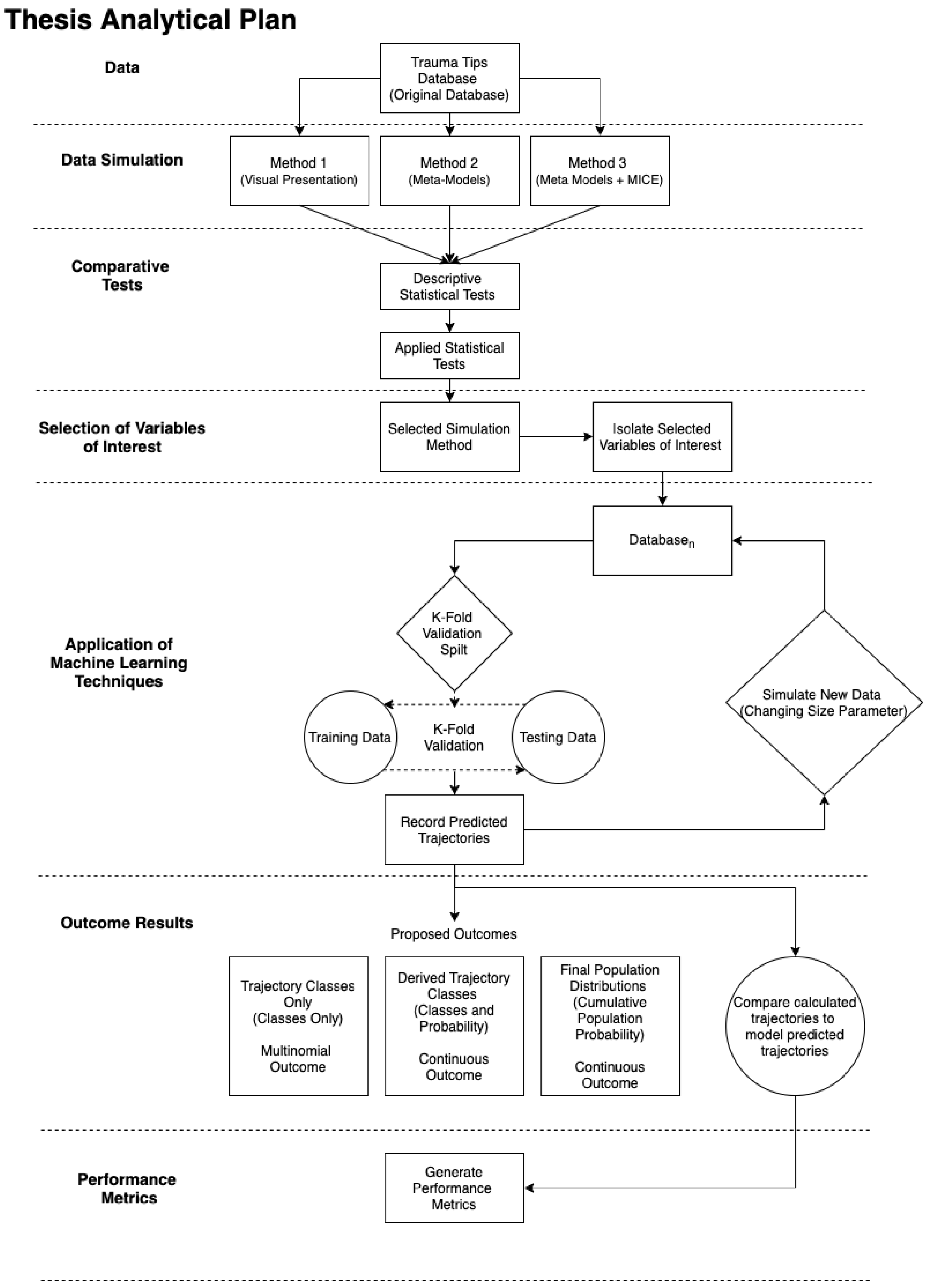
Thesis Analytical Plan
Data
Trauma Tips Database (Original Database)
Data Simulation
Method 1 (Visual Presentation)
Method 2 (Meta-Models)
Method 3 (Meta Models + MICE)
Comparative Tests
Descriptive Statistical Tests
Applied Statistical Tests
Selection of Variables of Interest
Selected Simulation Method
Isolate Selected Variables of Interest
Application of Machine Learning Techniques
Database_n
K-Fold Validation Spilt
Training Data
K-Fold Validation
Testing Data
Simulate New Data (Changing Size Parameter)
Record Predicted Trajectories
Outcome Results
Proposed Outcomes
Trajectory Classes Only (Classes Only) Multinomial Outcome
Derived Trajectory Classes (Classes and Probability) Continuous Outcome
Final Population Distributions (Cumulative Population Probability) Continuous Outcome
Compare calculated trajectories to model predicted trajectories
Performance Metrics
Generate Performance Metrics

Figure 1: Methodological Approach Flow Diagram

A) Reclassification of Variables

Qualitative variables were first reclassified into categorical variables, allowing their use as quantitative variables, making them more suitable for simulation and eventual application within the machine learning models. In particular, reclassification was applied to:

- Sport participation before the traumatic event. Reclassified using Mitchell, Haskell, Snell, Van Camp, & others (2005) classification of sports according to their static and dynamic components.
- Employment or Profession at the time of traumatic event. Reclassified using Centraal Bureau voor de Statistiek (CBS) employment classification (Fouarge, Dijksman, & others, 2015).
- Psychotropic medication usage at time of traumatic event, or historically. Classification of medication in line with their general pharmaceutical classification.
- Personal or family history of psychiatric conditions. Reclassified using major DSM-V (Association & others, 2013) themes.
- Relation of family members with noted psychiatric conditions. Reclassified under the broad categories of first, second or third degree relation.

B) Removal of Retrospective Questionnaires

For individual cases where self-report questionnaires were reported retrospectively, this data was removed as the reliability of the data from these questionnaires is considered low. The likelihood of retrospective reporting for the self-reported PTSD screening instruments was determined through calculating the reporting date, which if different from the specified time point would support their removal. Although this increased the level of missingness within the data, it was determined that retrospective reporting would have a greater impact upon the reliability of the results and therefore the later model than the induced missingness.

C) Time Based-Grouping

Finally, participants were regrouped according to specific time windows of the actual timing of their assessments in relation to their traumatic event. This temporal grouping was critical for the correct application of the bayesian latent growth mixture model formula, which determined trajectory membership from CAPS score (Zuiden et al., n.d.). Therefore, in line with these guidelines, CAPS scores were classified according to the following time points which define the time between traumatic event and CAPS assessment: T3a, 0-60 days; T3b, 61-136 days; T4, 137-273 days and T5, 274+ days.

Data Simulation

Three methods for the generation of the synthetic data are used to determine how the data should be simulated for this study. The first methods is the *conventional wisdom* (Skrondal, 2000), wherein variables are considered as independent, and thus simulated in isolation.

By contrast, the second and third methods use meta-models (Skrondal, 2000), assuming that observations are the product of functions between variables. Given the complexity of both the use data and the development of meta-models, the R Package OpenMX (Neale et al., 2016) is used to generate the underlying structural model and simulate data accordingly.

In the second model missing data is retained. Whilst the third model applies multiple imputation, through the R package MICE (Buuren & Groothuis-Oudshoorn, 2010), to address the issue of missing data. The application of multiple imputation, will allow the evaluation of the influence of missingness on the synthetic data.

As previously mentioned, data will be simulated using a variety of sample sizes. Specifically, these will be generated both in line with the existing data, which the synthetic data is based upon (n=852) as well as in line with smaller sample sizes (n=50) and considering large sample sizes (n=1000). The use of this range, and regular intervals accordingly, will as stated demonstrate the influence of sample size as well as enable additional conclusions to be drawn.

At each sample size, data sets will be simulated 100 times, to address the issue of potential bias in the process of simulation. Statistical tests will be applied to each individual data set before being averaged across all 100 tests forming an estimation of the true value, based on the population parameters.

Comparative Testing: For Data Simulation Methodology

To conclude which method of generating synthetic data both mimics and behaves similarity during statistical testing as the existing data, a series of comparative tests are applied. It is important to note that these tests are only used as a proof of principle, validating the performance of the synthetic data, with no optimization or interpretations made regarding the prognostic modeling from the selected synthetic data. As a result, only comparisons regarding similarity are made between the existing and synthetic data.

The planned comparative tests can be divided into two types: A) Descriptive Statistical Tests and B) Applied Statistical Tests. The first, will allow the comparison between the overall distribution of variables within existing and synthetic data sets, whereas the applied tests will confirm the similarities in behaviour of the tests when ML techniques are applied to the existing and synthetic data sets.

The simulation technique which demonstrates the highest similarity to the existing data, will be used to simulate data in the next part of the research.

A) Descriptive Statistical Tests

To examine whether the synthetic data sets are similar in their distributions, data will be taken for a subset of variables, and compared on traditional descriptive statistics (mean, standard deviation, skewness). Those which are most similar, as seen through comparative statistical testing, will be considered as closest to the existing data set.

B) Applied Statistical Tests

To examine how data behaves when statistical tests are applied to them, ML techniques are applied existing data, before applying the same model, using matching parameters to the synthetic data sets in turn. From this, the core measures of performance (sensitivity, specificity and Area Under the Curve) will be compared. Those synthetic data sets most similar to the existing data, will have performance measure scores which are similar (when averaged across the multiple iterations and sample sizes). Two techniques are applied, a decision tree, as these present a simple approach to classification, and Support Vector Machines (SVMs), given the large amount of literature supporting the use of SVMs in PTSD Prognosis (Galatzer-Levy et al., 2014; Ramos-Lima et al., 2020).

Selection of Variables of Interest

As this research provides a recommendation regarding the most appropriate statistical model to be used in the development of a trajectory focused screening instrument. The model should be based upon predictive variables which are available at the time of hospital admission. In the case of the existing data, these are self-reported measures collected at Time Points 1 and 2, and in particular those without the requirement of clinical assistance. These predictor variables can be categorized into 7 groups (Figure 2). This includes variables detailing a participants core demographic information (Age and Sex), their self-report measures (IES-R, SPAN, TSQ), their reported physical activity and employment information, information regarding the type and traumatic experience generally in addition to their physical and psychological health histories.

The developed model predicts a participants trajectory, as calculated by the bayesian latent growth mixture model (Zuiden et al., n.d.). These trajectories are determined using a participants CAPS score over time. Therefore, the clinical scores from the CAPS interviews are selected as parts of these variables of interest.

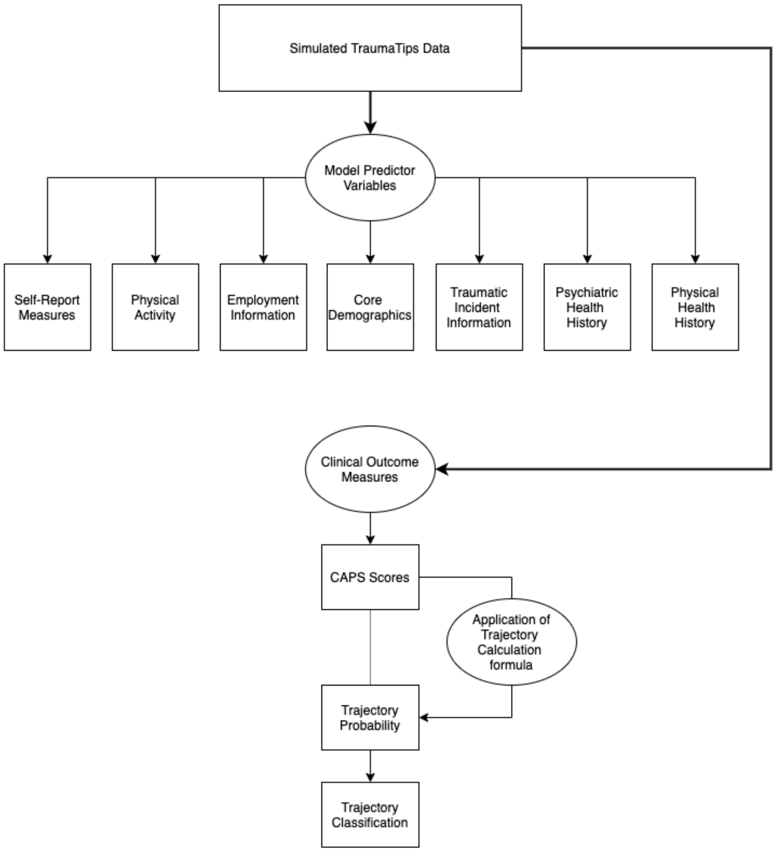
Simulated TraumaTips Data
Model Predictor Variables
Self-Report Measures
Physical Activity
Employment Information
Core Demographics
Traumatic Incident Information
Psychiatric Health History
Physical Health History
Clinical Outcome Measures
CAPS Scores
Application of Trajectory Calculation formula
Trajectory Probability
Trajectory Classification

Figure 2: Variables of Interest Summary

Outcome Results

Considering that a participants true trajectory can be considered as either a categorical trajectory classification, or numeric probability of trajectory membership. When developing the predictive model, three potential outcomes can be derived through ML techniques. Considering these different potential outcomes is important, with each providing clear advantages and disadvantages both methodological and through the potential application of model. These three potential outcome paths can be summarized as: A) Considering only the trajectory class, B) Deriving the trajectory class from the probability produced, and C) Comparing only the final distribution of the population, not comparing individual cases.

A) Trajectory Classes Only

This path would predict a participants trajectory only as a categorical classification. As a multinomial approach, in practice this would provide the simplest comparison through comparing the predicted categorical classification to that of the true value determined from the CAPS score.

Although having the most direct clinical application, providing a clear indication to the user which trajectory class they present in. This is problematic, as it presents with no margin of error between trajectory classes.

Meaning it is not possible to determine how close a participant is to other trajectory classes as only a single trajectory is provided.

B) Derived Trajectory Classes

In contrast to A, this path would predict the probability of trajectory membership, before deriving the trajectory class from the highest probability. In practice, this outcome path behaves most similarly to the calculation which derives the true trajectory, through deriving a trajectory class from the highest probability (Zuiden et al., n.d.).

In addition to having the direct clinical application through providing a clear indication to the user which trajectory class they present in. This path also overcome the issue regarding the margin of error present in A, as this allows the user, or associated clinical professional to see the probability of membership into other trajectory categories. However, this also presents some potential issues regarding those who score similarly between one or more classes, for example 33% vs 32%. The impact, and ways to address these similarities is a topic for future research as without specific examples of the likelihood and conditions this occurs, only speculations can be made.

Finally, some methodological concerns can also be raised regarding the calculation of Performance Metrics, given the comparison of continuous variables, through trajectory probability membership. Although these can still be easily calculated comparing the derived classes from the probabilities, further investigation should be made into the most effective methodology to compare these probabilities. However, the use of both categorical and continuous outcome variables, similar to those of the original trajectory calculation presents the highest degree of potential for future application.

C) Final Population Distribution

Finally, rather than comparing individuals within the population, this comparison would be made only for the distribution of individuals in each category. This cumulative approach, has the least amount of clinical application, however presents the greatest potential for demonstrating population distributions in PTSD. In practice, to derive these population distributions, either path A or B could be used, before distributions are calculated accordingly. Overall, although useful for future research, this is only useful when considering application alongside one of the aforementioned paths.

As a whole, these methods are being considered both individually and in combination, with a combination of B and C being presented as most useful through providing specific individual level detail, alongside the population wide distributions. However as specifying the exact outcome path would limit the ML techniques to only those producing a numeric result (no binary or multinomial). Until all the discussed techniques can be applied, pathway A should not be withdrawn as a suitable outcome result, with multinomial results presenting prominently in the wider prognosis research for PTSD (Ramos-Lima et al., 2020).

Application of Machine Learning Techniques

To make sufficient recommendations regarding the most appropriate ML technique for deriving a statistical model. A variety of different techniques are applied which are supported in previous research across PTSD prognosis research as well as both psychiatric and wider physical medical research. Alongside this, these techniques are selected based upon their outcome results, with classification techniques providing a nominal or binary outcome and regression techniques providing numeric, continuous outcomes. This, as highlighted in the Outcome Results section, is important in the way in which the the model recommendations are made due to the results produced. Table 1 summarizes the planned techniques, their R packages as well as their outcome measures.

Table 1: Summary of Machine Learning Techniques

Machine Learning Technique	**R Package**	**R Function**	**Outcome Result Type**
Support Vector Machines (SVMs)	e1071 (Meyer, Dimitriadou, Hornik, Weingessel, & Leisch, 2020)	*svm()*	Classification
Support Vector Regression (SVR)	e1071 (Meyer et al., 2020)	*svm()*	Regression
Regression (Multivariate)	stat (R Core Team, 2020)	*glm(), lm()*	Regression
Regression (Logistic)	stat (R Core Team, 2020)	*glm()*	Classification
K-Nearest Neighbour	class (Venables & Ripley, 2002)	*knn()*	Regression / Classification
Decision Trees (Recursive)	tree (Ripley, 2019)	*tree()*	Regression / Classification
Decision Trees (Random Forest)	randomForest (Liaw & Wiener, 2002)	*randomForest()*	Regression / Classification
Decision Trees (XGBoosted)	xgboost (Chen et al., 2021)	*xgb.train()*	Regression / Classification
Neural Networks	nnet (Venables & Ripley, 2002)	*nnet()*	Regression / Classification

These techniques are applied using a K-Fold cross validation (KCV) technique (Stone, 1974) to a range of different sized synthetic simulated data (n=50 to n=10000). The use of the KCV resampling technique, is completed to assist in the evaluation of the applied ML technique. Whereas the range of sample sizes, is used to investigate the influence it has upon key performance indicators (see Performance Metrics). Additionally, each ML technique will be applied across 100 different synthetic data sets of each size, to reduce the influence of random effects on the models generated.

Performance Metrics

The performance metrics of sensitivity, specificity and Area under the Curve (AUC) are used when discussing model performance. These metrics were selected primarily due to their presence in the field of clinical research (Dwyer, Falkai, & Koutsouleris, 2018), thus making this research accessible to other researchers. But also to draw effective comparisons between previous research which has examined PTSD prognosis specifically (Ramos-Lima et al., 2020). These metrics will be displayed visually (as seen in figure 3), displaying both the mean and a subset of individual simulations.

From this, several questions will be addressed:

- 1) Which technique has the highest overall metric (across all three measures) at the proposed Traumatips sample size (n=852).
- 2) Which technique has the highest overall metric, regardless of sample size.
- 3) Which technique has the highest overall metric below the proposed Traumatips sample size, and how does this change as the size increases?

Through addressing these specific questions, recommendations regarding which ML technique is most appropriate for the development of this statistical model can be made. These recommendations will provide the

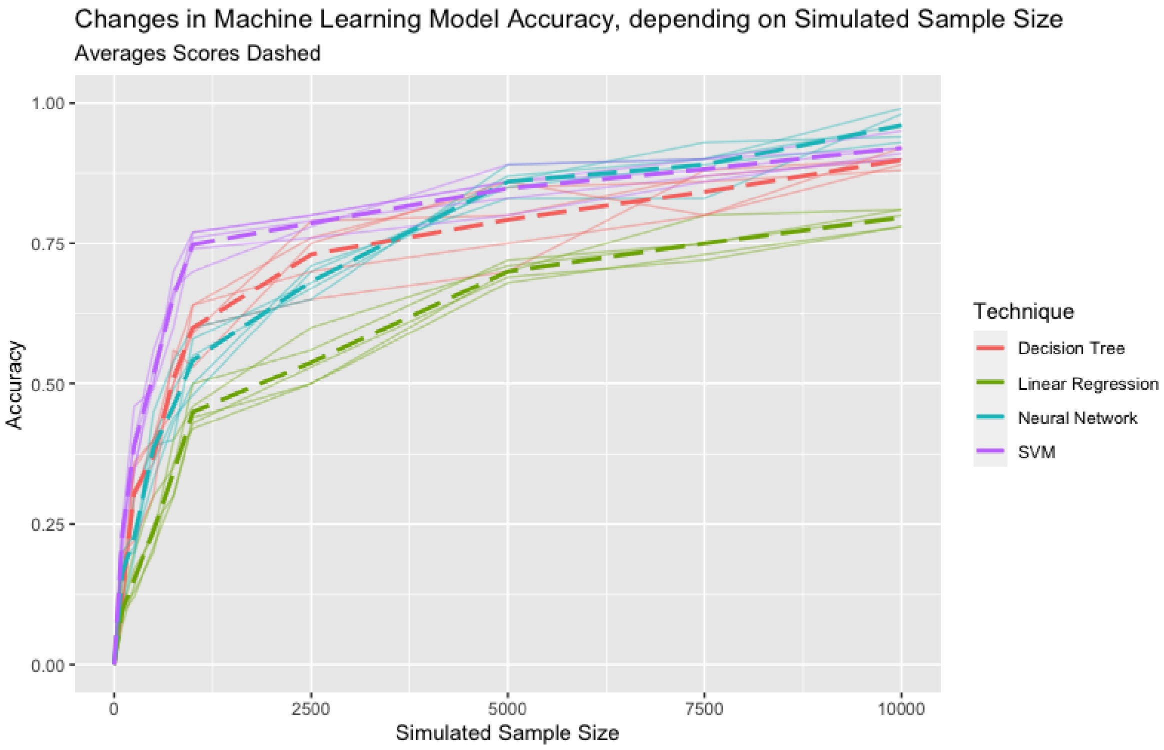

Figure 3: Mock Up Results Performance Metrics

foundation for the trajectory focused screening instrument to be developed from, using the Traumatips data set. Due to the scope of this research, it is likely that multiple methods will be recommended, however the potential advantages and disadvantages of these methods will be discussed in the recommendations made.

References

Ambler, G., Brady, A. R., & Royston, P. (2002). Simplifying a prognostic model: A simulation study based on clinical data. *Statistics in Medicine*, *21*(24), 3803–3822.

Association, A. P., & others. (2013). *Diagnostic and statistical manual of mental disorders (dsm-5)*. American Psychiatric Pub.

Augsburger, M., & Galatzer-Levy, I. R. (2020). Utilization of machine learning to test the impact of cognitive processing and emotion recognition on the development of ptsd following trauma exposure. *BMC Psychiatry*, *20*(1), 1–11.

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Gusman, F. D., Charney, D. S., & Keane, T. M. (1995). The development of a clinician-administered ptsd scale. *Journal of Traumatic Stress*, *8*(1), 75–90.

Bonanno, G. A. (2004). Loss, trauma, and human resilience: Have we underestimated the human capacity to thrive after extremely aversive events? *American Psychologist*, *59*(1), 20.

Bonanno, G. A., & Diminich, E. D. (2013). Annual research review: Positive adjustment to adversity–trajectories of minimal–impact resilience and emergent resilience. *Journal of Child Psychology and Psychiatry*, *54*(4), 378–401.

Breslau, N., Davis, G. C., Andreski, P., & Peterson, E. (1991). Traumatic events and posttraumatic stress disorder in an urban population of young adults. *Archives of General Psychiatry*, *48*(3), 216–222.

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, *25*(24), 4279–4292.

Buuren, S. van, & Groothuis-Oudshoorn, K. (2010). Mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 1–68.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., … Li, Y. (2021). *Xgboost: Extreme gradient boosting*. Retrieved from https://CRAN.R-project.org/package=xgboost

Cruz, J. A., & Wishart, D. S. (2006). Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics*, *2*, 117693510600200030.

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, *14*, 91–118.

Foa, E. B., Riggs, D. S., Dancu, C. V., & Rothbaum, B. O. (1993). Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *Journal of Traumatic Stress*, *6*(4), 459–473.

Fouarge, D., Dijksman, S., & others. (2015). *Beroepenindeling roa-cbs 2014 (brc 2014)*. Maastricht University, Research Centre for Education; the Labour Market (ROA).

Freedman, S. A., Eitan, R., & Weiniger, C. F. (2020). Interrupting traumatic memories in the emergency department: A randomized controlled pilot study. *European Journal of Psychotraumatology*, *11*(1), 1750170.

Galatzer-Levy, I. R., Huang, S. H., & Bonanno, G. A. (2018). Trajectories of resilience and dysfunction following potential trauma: A review and statistical evaluation. *Clinical Psychology Review*, *63*, 41–55.

Galatzer-Levy, I. R., Karstoft, K.-I., Statnikov, A., & Shalev, A. Y. (2014). Quantitative forecasting of ptsd from early trauma responses: A machine learning application. *Journal of Psychiatric Research*, *59*, 68–76.

Kearns, M. C., Ressler, K. J., Zatzick, D., & Rothbaum, B. O. (2012). Early interventions for ptsd: A review. *Depression and Anxiety*, *29*(10), 833–842.

Kononenko, I. (2001). Machine learning for medical diagnosis: History, state of the art and perspective. *Artificial Intelligence in Medicine*, *23*(1), 89–109.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, *13*, 8–17.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22. Retrieved from https://CRAN.R-project.org/doc/Rnews/

Marmer, C., & Weiss, D. (1997). The impact of event scale–revised. *Assessing Psychological Trauma and PTSD. Guilford, New York*, 399–411.

McLean, S. A., Ressler, K., Koenen, K. C., Neylan, T., Germine, L., Jovanovic, T., ... others. (2020). The aurora study: A longitudinal, multimodal library of brain biology and function after traumatic stress exposure. *Molecular Psychiatry*, *25*(2), 283–296.

Meltzer-Brody, S., Churchill, E., & Davidson, J. R. (1999). Derivation of the span, a brief diagnostic screening test for post-traumatic stress disorder. *Psychiatry Research*, *88*(1), 63–70.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2020). *E1071: Misc functions of the department of statistics, probability theory group (formerly: E1071), tu wien* (R package version 1.7-4).

Mitchell, J. H., Haskell, W., Snell, P., Van Camp, S. P., & others. (2005). Task force 8: Classification of sports. *Journal of the American College of Cardiology*, *45*(8), 1364–1367.

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*(11), 2074–2102.

Mouthaan, J., Sijbrandij, M., Reitsma, J. B., Gersons, B. P., & Olff, M. (2014). Comparing screening instruments to predict posttraumatic stress disorder. *PLoS One*, *9*(5), e97183.

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., ... Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, *81*(2), 535–549.

Passos, I. C., Mwangi, B., Cao, B., Hamilton, J. E., Wu, M.-J., Zhang, X. Y., ... others. (2016). Identifying a clinical signature of suicidality among patients with mood disorders: A pilot study using a machine learning approach. *Journal of Affective Disorders*, *193*, 109–116.

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, *380*(14), 1347–1358.

Ramos-Lima, L. F., Waikamp, V., Antonelli-Salgado, T., Passos, I. C., & Freitas, L. H. M. (2020). The use of machine learning techniques in trauma-related disorders: A systematic review. *Journal of Psychiatric Research*, *121*, 159–172.

Ranganathan, P., Pramesh, C., & Buyse, M. (2016). Common pitfalls in statistical analysis: The perils of multiple testing. *Perspectives in Clinical Research*, *7*(2), 106.

R Core Team. (2020). *R: A language and environment for statistical computing.* Retrieved from https://www.R-project.org

Ripley, B. (2019). *Tree: Classification and regression trees.* Retrieved from https://CRAN.R-project.org/package=tree

Schoot, R. van de, Sijbrandij, M., Depaoli, S., Winter, S. D., Olff, M., & Van Loey, N. E. (2018). Bayesian ptsd-trajectory analysis with informed priors based on a systematic literature search and expert elicitation. *Multivariate Behavioral Research*, *53*(2), 267–291.

Shalev, A. Y., Gevonden, M., Ratanatharathorn, A., Laska, E., Van Der Mei, W. F., Qi, W., ... others. (2019). Estimating the risk of ptsd in recent trauma survivors: Results of the international consortium to predict ptsd (icpp). *World Psychiatry*, *18*(1), 77–87.

Short, N. A., Morabito, D. M., & Gilmore, A. K. (2020). Secondary prevention for posttraumatic stress and related symptoms among women whohave experienced a recent sexual assault: A systematic review and meta-analysis. *Depression and Anxiety.*

Skrondal, A. (2000). Design and analysis of monte carlo experiments: Attacking the conventional wisdom. *Multivariate Behavioral Research*, *35*(2), 137–167.

Smid, G. E., Mooren, T. T., Mast, R. C. van der, Gersons, B. P., & Kleber, R. J. (2009). Delayed posttraumatic stress disorder: Systematic review, meta-analysis, and meta-regression analysis of prospective studies. *Journal of Clinical Psychiatry*, *70*(11), 1572.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, *36*(2), 111–133.

Taylor, T. R., Ford, D. N., & Ford, A. (2010). Improving model understanding using statistical screening. *System Dynamics Review*, *26*(1), 73–87.

Varoquaux, G. (2018). Cross-validation failure: Small sample sizes lead to large error bars. *Neuroimage*, *180*, 68–77.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth). Retrieved from http://www.stats.ox.ac.uk/pub/MASS4

Webb, C. A., Cohen, Z. D., Beard, C., Forgeard, M., Peckham, A. D., & Björgvinsson, T. (2020). Personalized prognostic prediction of treatment outcome for depressed patients in a naturalistic psychiatric hospital setting: A comparison of machine learning approaches. *Journal of Consulting and Clinical Psychology*, *88*(1), 25.

Zlotnick, C., Davidson, J., Shea, M. T., & Pearlstein, T. (1996). Validation of the davidson trauma scale in a sample of survivors of childhood sexual abuse. *Journal of Nervous and Mental Disease*.

Zuiden, M. van, Engel, S., Karchoud, J., Mouthaan, J., Wise, T., Sijbrandij, E. M., … Schoot, R. van de. (n.d.). *A bayesian investigation of sex-specific longitudinal trajectories of ptsd symptoms following traumatic injury.*