A Traumatic Brain Injury Prescreening Tool for Intimate Partner Violence Patients Using Initial Clinical Reports and Machine Learning

Abheet Singh Sachdeva[1], Avery Bell[2], Dr. Jacob Furst, PhD[3], Dorothy A. Kozlowski, PhD[3], Sonya Crabtree-Nelson, PhD, LCSW[3], Daniela Raicu, PhD[3]

[1]Baylor University, Waco, Texas; [2]Northern Arizona University, Flagstaff, Arizona; [3]DePaul University, Chicago, Illinois

Abstract

Research studies have presented an unappreciated relationship between intimate partner violence (IPV) survivors and symptoms of traumatic brain injuries (TBI). Within these IPV survivors, resulting TBIs are not always identified during emergency room visits. This demonstrates a need for a prescreening tool that identifies IPV survivors who should receive TBI screening. We present a model that measures similarities to clinical reports for confirmed TBI cases to identify whether a patient should be screened for TBI. This is done through an ensemble of three supervised learning classifiers which work in two distinct feature spaces. Individual classifiers are trained on clinical reports and then used to create an ensemble that needs only one positive label to indicate a patient should be screened for TBI.

Introduction

Intimate Partner Violence and Traumatic Brain Injuries

Intimate partner violence (IPV) can be defined as behavior that "causes physical, sexual or psychological harm, including physical aggression, sexual coercion, psychological abuse and controlling behaviors" in an intimate relationship.[1] This sort of harm is very common with one in three women and one in four men having experienced severe physical IPV at least once in their lifetime.[2] It is estimated that between 35%-90% of IPV survivors have experienced at least one head related injury.[3] These head injuries can be broadly classified as traumatic brain injury (TBI). TBI can be described as an "alteration in brain function caused by external force."[4] Symptoms of TBI include altered mental state, loss of consciousness, and post-traumatic amnesia.[5] Loss of consciousness is regarded as an important symptom of TBI, but is not present in all brain injury cases.[6] TBI resulting from IPV is an injury that is often underreported.[7] This is due to IPV being underreported, and as a result, IPV induced brain injuries remain undetected. When presenting to the emergency department, some studies have found that 72% of domestic violence victims were not identified due to a lack of visible external injuries.[8] Similarly, IPV related TBI is estimated to be 11-12 times greater than the published incidence for other forms of TBI.[9] A vast majority of the literature on IPV is centered around women, with the data related to men being minimal.[10] Thus, the field has demonstrated a need for a solution to the underreporting of IPV, IPV related TBI, and representation of men in datasets.

IPV-TBI Screening Tools

The World Health Organization has validated several screening tools to identify TBIs, but none of these tools have been adapted to screen for TBI in the context of IPV.[3] Validated TBI screening tools phrase their survey questions in the context of the situation where a brain injury might have occurred. A modified version of the Brain Injury Screening Questionnaire (BISQ) has been proposed to screen for TBI in the context of IPV. Initial testing of BISQ-IPV indicates that screening in the context of IPV reveals additional brain injuries when compared to BISQ.[11] This initial testing has not been validated and so as mentioned before, it cannot be used as a validated screening approach. Similarly a modified version of the HELPS screening tool was used to estimate how often women were at risk for a brain injury.[12] This method provided a stringent criteria for identifying brain injury by asking about blows to the head, treatments received at the emergency room, loss of consciousness, and problems related to head injuries.[12] Similar to the BISQ-IPV tool it is not a validated screening tool. The Boston Assessment of TBI-Lifetime (BAT-L), a validated screening tool used to identify lifetime TBI in post 9/11 veterans, was adapted to IPV patients.[13] Its results were compared to a well-validated Ohio State University TBI Identification Method (OSU-TBI-ID) and results indicated good performance.[13] This screening tool relies on a forensic approach that requires a patient to remember the events of a brain injury. Given that a symptom of brain injuries is posttraumatic amnesia they may not remember the event, thereby impacting the screening's results. As mentioned before, patients are not always likely to report their symptoms, and a screening tool heavily reliant on chronological order and differentiation between symptom etiology may not be entirely useful. A tool called CHATS from the Ohio Domestic Violence Network is in the process of being validated.[14]

Electronic Health Record Analysis

Diagnosis or identification of diseases through the use of clinical text has been done in other medical disciplines. One study identified keywords associated with an electronic health record (EHR) to discern patients at risk of HIV.[15] This study reduced a list of terms with high Term Frequency-Inverse Document Frequency (TF-IDF) scores through univariate chi-square testing to create a set of statistically significant keywords.[15] A manual selection from this set created the keyword list they used in their predictive model.[15] The key words identified include "hiv", "homosexual", and "tested".[15] Using key words derived from EHR to assess HIV risk is not insightful when one of the words identified to be associated with high risk is 'hiv', indicating the healthcare professional has already identified the disease itself. Another study designed a custom dictionary to extract terms relevant to schizophrenia in a set of clinical notes.[16] This was done by building a matrix that indicates presence or negation of terms and then using Latent Dirichlet Allocation (LDA) to identify topics and reduce features.[16] The final selection was done based on the topic weights.[16] Although this methodology has a heavier reliance on statistical correlations, the act of manual selection reduces the accuracy of the statistics and can potentially lead it to be less relevant. Finally, another study used a manual dictionary of phrases related to cognitive decline to identify symptoms of mild cognitive impairment in order to train a prediction model.[17] The manual implementation each of these studies used relies heavily on the expertise of the individual and is contingent on each relevant word being identified, with respect to spelling errors, synonyms, and alternate phrases. Similarly, if additional EHR were to be added to this dataset, the same manual process would need to be repeated for each new record. Manual dictionaries have many downfalls in their inability to be generalized. Although they, like black box methods, can yield high results on a specific dataset, it is difficult to apply the same methodology to new cases or continually identify risk.

As stated before, methodologies reliant on manual dictionaries are difficult to generalize to new texts. The following studies have used a non-manual creation of dictionaries or do not use a dictionary at all in their identification of labels. One study used vectorized clinical notes and clustering to find distinguishing characteristics of heart disease EHR.[18] The clustering approach makes the method unsupervised, and as stated by its limitations its avoidance of diagnosis codes makes the labels descriptive rather than definitive. An alternative to dictionaries is the use of Word2Vec and a bag-of-words approach to generate ICD-9 codes related to rheumatology from EHRs.[19] EHR data is dependent on health professional investigation and so in cases where documentation is sparse, extracted ICD-9 codes may be incomplete or inaccurate.[19] Another study used large-scale support vector machine (SVM) based classifiers to extract a diagnosis status from intensive care unit clinical reports.[20] This provides some direction as to a method by which diagnosis can occur without dictionary abstraction, and is generalizable. It is, however, using only one classifier which is subject to overfitting and the volatility of the notes.[20] There is also an example in which black box modeling in the form of an artificial neural network was used to identify clinically relevant TBI cases in children through computed tomography and demographic data.[21] As is with most black boxes, high accuracies are achievable, but understanding how it is done is not. Thus, to the best of our knowledge TBI diagnosis through clinical text specifically in the case of IPV patients has not been done. However, work has been done that analyzes electronic health records to establish health effects and key associations between IPV and TBI.[22] This methodology revolved around extracting clinical terms from EHRs to establish a relationship with IPV and TBI. [22] This analysis revealed that IPV induced TBI has a relationship with other acute conditions including concussion, chronic post-traumatic headache, hematoma, and delirium.[22]

Existing literature indicates free text clinical notes can be used to identify illnesses. There are existing TBI screening tools that can be used to identify TBI in the context of IPV, despite not being validated. IPV induced TBI symptoms are often difficult to identify, and so healthcare professionals do not always have all the information they need to diagnose a TBI. In addition, during an Emergency Department visit related to an IPV incident, many dimensions of the situation are being addressed, including immediate safety, the need to find shelter, and sometimes law enforcement. We aim to create a way to flag cases, to prioritize screening for TBI by employing an ensemble of multiple supervised learning classifiers trained on clinical IPV reports.

Methods

Dataset

The dataset was acquired from the Emergency Department of an urban Midwest hospital and represents patients from June 2017 through June 2021. It was collected and analyzed in accordance with DePaul University and the hospital's IRB approval. The last approved date was 11/29/2023. Figure 1 demonstrates the demographic breakdown of this dataset. The median and mean age of the patients are 34 and 36.7 years old respectively, and the distribution of these ages can be seen in Figure 1a. There are 522 Female patients and 162 Male patients, as seen in Figure 1b. The racial breakdown of these patients is 225 Latino, 224 White, 98 African American, 75 Asian, 55 Other, 5 Native American,

and 2 Not Reported. Each patient is an IPV survivor and was designated as such by the hospital. This dataset also contains an Initial Clinical Reports section, that includes the first set of clinical notes for patients who have been identified as IPV survivors, and a TBI Reports section, which contains an additional set of clinical notes that indicate an IPV survivor disclosed a head injury.

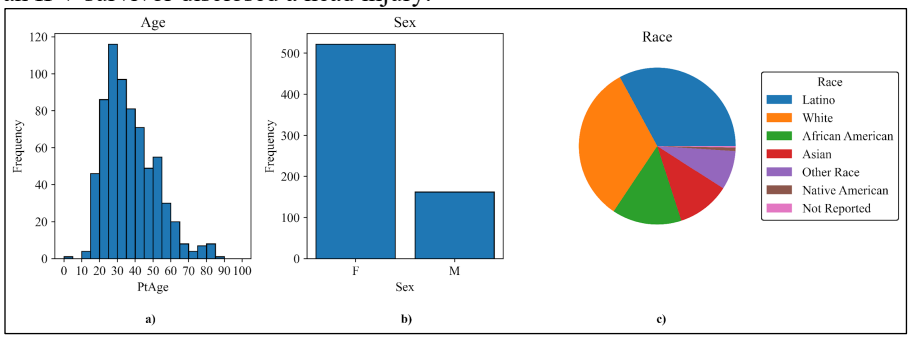

Figure 1. Demographics of patients represented in the entire IPV-TBI dataset. a) The distribution of patient ages. b) Comparison of sexes represented. c) Percentage breakdown of races represented.

Ground Truth Cases

To utilize existing supervised learning classification methods, Ground Truth cases, or cases that definitively contain or do not contain a TBI, need to be identified. Of the 564 reports represented in the TBI Reports, 71 patients have an ICD-10 diagnosis code related to TBI. These cases have been designated as Ground Truth Positive cases. Of the 686 reports represented in the Initial Clinical Reports section, 122 patients do not have a corresponding entry in the TBI Report section, indicating that the patient did not disclose injury to the face, neck, or head in connection with the IPV incident. These cases have been designated as Ground Truth Negative cases. A manual review strongly suggests that these Ground Truth cases are likely representative and can thus be considered as ground truth. The remaining 493 patients have an Initial Clinical Report and TBI report, but do not have an associated ICD-10 code to indicate an injury to the head, neck or face occurred. Therefore, they cannot be assigned a Ground Truth label, and as a result, were omitted from our study.

Table 1: Breakdown of the Reports and number of cases.

Label	Initial Clinical Report	TBI Report
Ground Truth Positive	71	71
Ground Truth Negative	122	0
Omitted	493	493
Total Cases	686	564

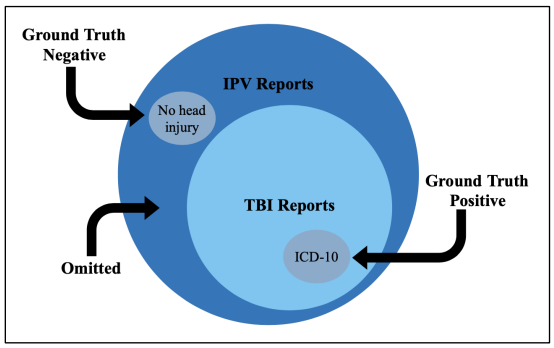

Figure 2: Breakdown of Ground Truth case origin.

Text Selection

The goal of this project was to create a model that identifies IPV survivors who should receive additional screening for TBI. Our model accomplishes this by measuring how similar a hospital clinical report is to reports with Ground Truth Positive and Negative TBIs. We chose to train our model on the earliest set of clinical notes taken upon arrival to the emergency room so that an IPV survivor at risk of TBI can be identified for screening as soon as an initial report is made. Thus, we chose to train our model on only the Initial Clinical Reports.

Clinical Note Preprocessing

We followed standard text preprocessing to prepare our clinical notes for analysis. Figure 3 uses a fictional example to delineate the preprocessing steps we used to prepare the Initial Clinical Reports for analysis. The preprocessing steps followed were stopword removal, lemmatization, and negation.

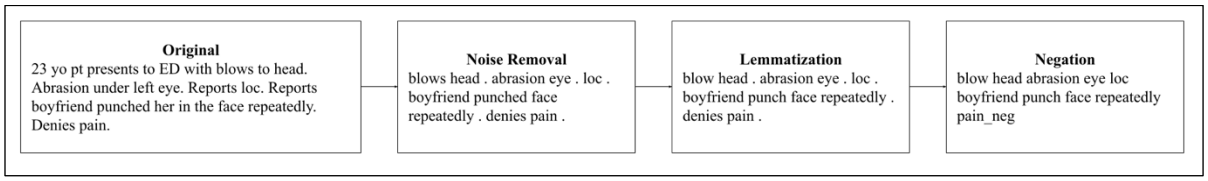

Figure 3. The text preprocessing steps displayed on a fictional report.

First, stopwords were removed from each individual clinical note. To do so, the documents were lowercased and tokenized using the spaCy library.[17,20,23] Stopwords are common words that don't contain much meaning to a computer, like conjunctions, articles, and pronouns.[23] The standard English list found in the nltk library set the foundation for the words that were to be removed.[24,25] However, in order to add words to the list that were more specific to this set of text, we used an iterative process. [15] This iterative process looked like the following. The top 30 most frequent words for the entire column were outputted and any word from this list not deemed to be important to meaning was manually added to the list. This was done so until the top 30 tokens contained only relevant words. These words included 'patient', 'states', 'left', 'right', and 'police'. Any token that was an abbreviation for a word removed based on the top 30 list, such as 'pt', 'l', and 'r', was also removed. Although important to physician documentation, the words we removed are not relevant to the computer model and instead add noise.[23]

Domain knowledge in the form of current screening procedures indicates that loss of consciousness is critical to TBI identification and thus steps were taken to standardize the phrase "lose consciousness."[3] To ensure that the replacement process accounted for any spelling errors as well as alternate phrasings, we used a combination of regular expressions and fuzz from the fuzzywuzzy library.[26,27] The final aspect of this step is the removal of non-alpha characters. We used regular expressions to remove all punctuation (excluding periods), numeric, and special characters from the text. [15,19,20]

Next, we lemmatized each note. Lemmatization reduces words to their basic forms, like changing 'running' to 'run'. This retains root word meaning while reducing the number of unique words that exist within a report. We first lemmatized a word as a noun and then lemmatized the resulting word as a verb. This approach allowed us to bypass part of speech tagging, a common procedure that did not perform well on smaller text test sets. Although this version fails to account for words whose lemmatized noun form is now a verb, their rarity makes the impact minimal and one that can be sustained by this dataset.

Last, we identified negated words. Words preceded by 'no', 'not' or 'deny' were tagged to indicate its meaning has changed. This was done so by appending a negation tag of '_neg' to the end of each token until the presence of a '.' token. Thereby resulting in the computer interpreting each word after any of the three words until the end of the sentence to be negated. Although this caused words that were not in fact being negated to indicate that they were, it allowed for lists of symptoms to all be negated at once. When tested on smaller sets of text, alternative options failed to add negation tags to all symptoms in a list after a negator word. This tradeoff was also deemed to be sustainable and had minimal impact. After performing these preprocessing steps, we translated the text into numeric values that can be used in standard machine learning techniques.

Individual Classifiers

The individual classifiers are Random Forest[28], fastText[29], and a centroid based classifer. We chose a three-classifier ensemble to explore different methods of prediction aggregation, and to represent different feature spaces in individual predictions. Three classifiers also represented the smallest number for which a majority could be achieved. The centroid based and Random Forest classifiers were trained in TF-IDF space. Term frequency refers to how often a word appears within a document, and document frequency refers to how many documents the word appears in.[30] Thus, a word has a high TF-IDF score if the word that appears many times in a document, and also does not appear frequently in other documents in a dataset.[30] TF-IDF scores can be thought of as word importance scores. Using TF-IDF representations of the Initial Clinical Reports in classification models allows us to assess which documents have the same sets of words with similar importance scores. A classification from a model trained in the TF-IDF space indicates which set of Ground Truth cases a given report is more similar to. Thus, a positive classification from our model would indicate that the report being evaluated was similar enough to the Ground Truth positive set of Initial Clinical reports to suggest screening the IPV survivor for TBI. This is the basis for our claim that the ensemble can be used as a prescreening tool for TBI evaluation.

The fastText model was selected because it trains in a semantic space. Several options exist, namely Word2vec, fastText, and ClinicalBERT. Due to the nature of clinical report text and its tendency to contain abbreviations, medical terms, and other words not found in a dictionary, picking a model that could handle this was important. fastText is the

best of these three options in this regard as it creates its vocabulary vector based on the words it encounters within the text it is trained on. In addition, clinical reports generally have some order in which the medical professional takes notes, but the context of a sentence before or after is not standardized and so relying on a model that needs a context window like Word2vec could prove unreliable. Lastly, the simplicity of the fastText training and testing compared to ClinicalBERT allowed us to choose a method that maintained the simplest route possible.

The Random Forest was initialized to the same random state for each run and contained 100 decision trees. Aside from initial hyperparameter decisions, no further optimization was done. fastText had three parameters which we tuned. The first is the epoch which was set to eight. This value was the highest value seen after conducting several grid searches on a subset of Ground Truth cases. It is important to note that fastText is a blackbox and utilizes randomness in its training process, thus the ideal parameters vary slightly. The second was the learning rate which was set to 0.9. Contrary to the selection of the epoch, this was the lowest learning rate seen during the grid search process. Finally, the wordNgrams was set to one. Throughout each grid search this parameter never changed from one and so this consistency indicated that it was the best value.

Aggregation Procedure and Selection

We explored two methods of aggregating the results of our individual classifiers. The first was named 'Any evidence is enough evidence', or 'Any Evidence' for short. In this method of aggregation, a final positive prediction was assigned if at least one individual classifier produced a positive prediction. The only scenario where the 'Any evidence' method assigns a negative prediction is when all three classifiers produce a negative prediction. The second is named 'Majority vote'. In this case, a final prediction is assigned based on agreement between two or more individual classifiers. There is one instance in which these two aggregation models disagree: two negative labels and one positive label. This difference is what results in the differing results of the two classifiers.

We assessed both methods of aggregation, as well as individual classifier performance, by utilizing 30 train/test splits. These splits followed an 80:20 train/test ratio. For each split, we trained our individual classifiers, used them to predict the test set, and used both methods of aggregation to acquire two sets of final ensemble predictions. The model performance metrics accuracy and sensitivity were calculated, and statistical testing was used to determine which method of aggregation performed better. The better method of aggregation was the model with the highest sensitivity was selected as the final aggregation method for our ensemble model.

As a part of training our individual classifiers, the train set was split into another 80:20 split, referred to as the classifier train and test sets. For 30 splits, each classifier was trained on the classifier train set and then tested on the classifier test set. From the 30 classifiers trained during the 30-classifier train/test splits, the final version selected was the model that performed closest to the average testing accuracy across all splits. This final model for each individual classifier was used as part of the ensemble to predict the initial test set.

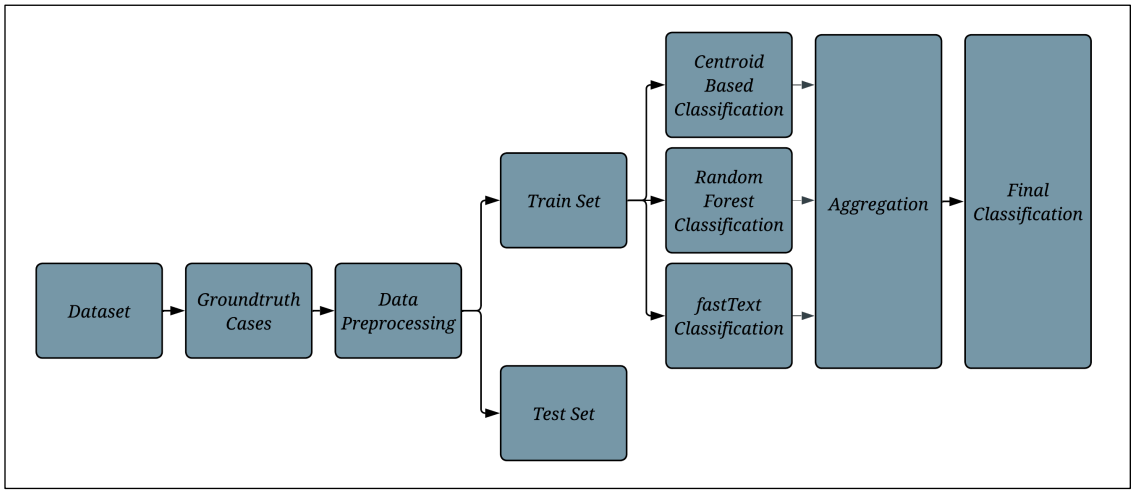

Figure 4: The methods displayed as a workflow diagram.

Results

Ground Truth Prediction

For each train/test split, accuracy, sensitivity, and specificity were calculated for each of the three individual classifiers and the two aggregated models. The calculated mean accuracy, sensitivity, and specificity for the individual and

aggregate classifiers along with their confidence intervals are presented in Table 2. Because the distributions of performance metrics were not normally distributed, bootstrapped confidence intervals were constructed.

Table 2: Performance of Ground Truth prediction using individual and aggregate classifiers with their 95% confidence intervals.

Model	**Accuracy**	**Sensitivity**	**Specificity**
Any Evidence Ensemble	0.911 [0.900, 0.920]	**0.940 [0.920, 0.960]**	0.895 [0.880, 0.910]
Majority Vote Ensemble	0.928 [0.920, 0.940]	0.914 [0.890, 0.930]	0.936 [0.920, 0.950]
Centroid Based	0.932 [0.920, 0.940]	0.912 [0.890, 0.930]	0.943 [0.930, 0.960]
Random Forest	0.927 [0.910, 0.940]	0.888 [0.860, 0.910]	0.949 [0.940, 0.960]
fastText	0.912 [0.900, 0.920]	0.907 [0.890, 0.930]	0.915 [0.900, 0.930]

Comparing Individual and Aggregation Classifiers

The performance of our individual classifiers compared to both ensembles is similar. To determine which model should be retained as a TBI prescreening tool for IPV survivors, we considered the real-world application of our model. It is better to screen a patient for a TBI when they do not have a brain injury, than to not screen a patient when they do have a brain injury. This leads us to prefer a model with a tendency towards false positives over false negatives. Because of this, we used a one tailed Wilcoxon signed-rank test to determine if the Any Evidence Ensemble, which has the highest mean sensitivity, performs statistically significantly better than the other classifiers.

Table 3 shows the pairwise test results comparing model sensitivities or i to the Any Evidence Ensemble. The null hypothesis of these tests is that mean sensitivity for the Any Evidence Ensemble is the same as i's mean sensitivity. The alternative hypothesis is that the mean sensitivity for the Any Evidence Ensemble is greater than i's mean sensitivity. A Bonferroni correction for multiple hypothesis testing was utilized, making the level that determines statistical significance to be $\boldsymbol{\alpha} = 0.0125$.

Table 3. Results for a one tailed Wilcoxon Test to determine if the Any Evidence Ensemble performs statistically significantly better with respect to sensitivity, $\boldsymbol{\alpha} = 0.0125$.

Model	**p-value**
Majority Vote Ensemble	0.002118
Centroid Based	0.001267
Random Forest	0.000076
fastText	0.001725

The results of the pairwise hypothesis tests indicate that the Any Evidence Ensemble has a statistically significantly higher sensitivity. Because of this, we selected the Any Evidence Ensemble as the model we present as a TBI prescreening tool for IPV survivors.

Discussion

In this work we used clinical reports from a set of IPV patients to create an aggregate classifier that identified patients who should be screened for TBI. The dataset consisted of 122 Ground Truth Negative, 71 Ground Truth Positive, and 493 Unlabeled reports. In a healthcare domain, it is not uncommon that sensitivity is prioritized over accuracy. This is because the impact of a falsely identified positive patient is an extra screening, whereas a falsely identified negative patient results in a TBI going undetected. Because of this, we selected a model that has the highest sensitivity. The chosen model, the Any Evidence Ensemble, had an accuracy of 0.911 [0.900, 0.920] and a statistically significantly higher sensitivity of 0.940 [0.920, 0.960]. In hospitals with a busy emergency department where staff resources are stretched, our ensemble would help prioritize patients who should receive additional screening. In hospitals looking to address TBI underdiagnosis in IPV patients, our ensemble would serve as a prescreening tool to identify TBI cases likely to be overlooked. The ensemble in better identifying TBI cases from the same reported data is able to work towards the improvement of the patient experience. We believe that this methodology is generalizable to other areas

of healthcare. Any field that generates clinical reports can measure report similarities to identify patients who are at risk for an underreported diagnosis. The application of this methodology in diverse healthcare disciplines can improve patient outcomes, increase identification of understudied phenomena, and prioritize hospital resources.

We propose our model be used in conjunction with a validated TBI screening tool to address TBI underdiagnosis in IPV survivors. Our ensemble identifies IPV survivors who should be screened for a TBI by measuring similarities between their Initial Clinical Report and Initial Clinical Reports with a TBI related diagnosis. After initial text preprocessing, we translate the Initial Clinical Report into a numeric representation using TF-IDF scores. TF-IDF scores can be thought of as importance scores for each word in a report. The centroid based and Random Forest classifiers were trained in a TF-IDF space. This means that they measure similarities between reports by identifying sets of words with similar importance scores. It is the presence of multiple words with similar importance scores that allows our classifiers to make predictions, and as such, an exploration into individual words that carry the most weight in our models is not insightful.

Limitations

Our ensemble is trained on Initial Clinical Reports for males and females. The current body of literature surrounding IPV and TBI often omits males from their dataset as they represent a minority of those who disclose IPV. It is possible that differing attitudes towards males who experience IPV results in clinical report note differences, however, no investigation into this was performed.

Clinical reports reflect a patient's willingness to disclose injuries, information a healthcare professional deems relevant, and internalized attitudes about IPV. As the negative stigma associated with IPV lessens, our ensemble will need to be retrained on clinical notes that reflect the current IPV attitudes to ensure it can correctly measure report similarities. Despite this, the ensemble's ability to improve on identification of TBI from the present reports is significant.

Our text preprocessing used a crude form of negation and lemmatization that did not rely on any form of part of speech (POS) tagging. Although this introduced some amount of noise into our resulting text, the impact is negligible. As more rigorous forms of POS tagging become available, integrating them into our methodology would refine our preprocessing steps.

Conclusion

We trained an ensemble of classification models on Initial Clinical Reports to create a prescreening tool that can be used to identify IPV survivors who should be screened for TBI. Our ensemble was selected because it had the highest sensitivity, or highest true positive rate. We believe the application of this tool will address the understudied relationship between IPV survivors and their resulting TBI. Further work will include assessing reporting differences between male and female IPV survivors, retraining the ensemble on updated clinical reports, and improving text preprocessing. In this study, we proposed an aggregated classification model that measures clinical report similarities on the earliest set of clinical notes to identify IPV survivors who should be screened for TBI. This methodology to prescreen patients from initial clinical reports for often overlooked diagnoses can be applied in other healthcare contexts.

References

1. Violence against women. [Internet]. [cited 2023 Aug 14]. Available from: https://www.who.int/news-room/fact-sheets/detail/violence-against-women
2. CDC. Fast facts: preventing intimate partner violence |violence prevention|injury center|CDC. [Internet]. 2022 [cited 2023 Aug 15]. Available from: https://www.cdc.gov/violenceprevention/intimatepartnerviolence/fastfact.html
3. Haag H (Lin), Jones D, Joseph T, Colantonio A. Battered and brain injured: traumatic brain injury among women survivors of intimate partner violence—a scoping review. Trauma, Violence, & Abuse. 2022 Oct 1;23(4):1270–87.
4. Brain injury overview. [Internet]. Brain Injury Association of America. [cited 2023 Aug 14]. Available from: https://www.biausa.org/brain-injury/about-brain-injury/basics/overview
5. Fortier CB, Kim S, Alexandra K. Assessment of IPV-related brain injury: what do we know and where do we go? Brain Injury professional. 2023 Jan;19(3):14–7.
6. Valera EM, Berenbaum H. Brain injury in battered women. J Consult Clin Psychol. 2003 Aug;71(4):797–804.
7. St. Ivany A, Schminkey D. Intimate partner violence and traumatic brain injury: state of the science and next steps. Family & Community Health. 2016 Jun;39(2):129.
8. Costello K, Greenwald BD. Update on domestic violence and traumatic brain injury: a narrative review. Brain Sciences. 2022 Jan;12(1):122.

9. Lifshitz J, Crabtree-Nelson S, Kozlowski DA. Traumatic brain injury in victims of domestic violence. Journal of Aggression, Maltreatment & Trauma. 2019 Jul 3;28(6):655–9.
10. Valera E. Intimate partner violence and brain injury: a selective overview. Brain Injury Professional. 2023 Jan;19(3):8–10.
11. Dams-O'Connor K, Bulas A, Lin Haag H, Spielman LA, Fernandez A, Frederick-Hawley L, et al. Screening for brain injury sustained in the context of intimate partner violence (IPV): measure development and preliminary utility of the brain injury screening questionnaire IPV module. J Neurotrauma. 2023 Apr 27;
12. Rajaram SS, Reisher P, Garlinghouse M, Chiou KS, Higgins KD, New-Aaron M, et al. Intimate partner violence and brain injury screening. Violence Against Women. 2021 Aug 1;27(10):1548–65.
13. Fortier CB, Beck BM, Werner KB, Iverson KM, Kim S, Currao A, et al. The Boston assessment of traumatic brain injury-lifetime semistructured interview for assessment of TBI and subconcussive injury among female survivors of intimate partner violence: evidence of research utility and validity. J Head Trauma Rehabil. 2022 Jun 1;37(3):E175–85.
14. Ohio Domestic Violence Network. Brain injury. [Internet]. [cited 2023 Aug 15]. Available from: https://www.odvn.org/brain-injury/
15. Feller DJ, Zucker J, Yin MT, Gordon P, Elhadad N. Using clinical notes and natural language processing for automated HIV risk assessment. J Acquir Immune Defic Syndr. 2018 Feb 1;77(2):160–6.
16. Liu Q, Woo M, Zou X, Champaneria A, Lau C, Mubbashar MI, et al. Symptom-based patient stratification in mental illness using clinical notes. J Biomed Inform. 2019 Oct;98:103274.
17. Penfold RB, Carrell DS, Cronkite DJ, Pabiniak C, Dodd T, Glass AM, et al. Development of a machine learning model to predict mild cognitive impairment using natural language processing in the absence of screening. BMC Med Inform Decis Mak. 2022 May 1;22(1):129.
18. Nagamine T, Gillette B, Pakhomov A, Kahoun J, Mayer H, Burghaus R, et al. Multiscale classification of heart failure phenotypes by unsupervised clustering of unstructured electronic medical record data. Sci Rep. 2020 Dec 7;10(1):21340.
19. Turner CA, Jacobs AD, Marques CK, Oates JC, Kamen DL, Anderson PE, et al. Word2Vec inversion and traditional text classifiers for phenotyping lupus. BMC Med Inform Decis Mak. 2017 Aug 1;17(1):126.
20. Marafino BJ, Davies JM, Bardach NS, Dean ML, Dudley RA. N-gram support vector machines for scalable procedure and diagnosis classification, with applications to clinical free text data from the intensive care unit. J Am Med Inform Assoc. 2014;21(5):871–5.
21. Ellethy H, Chandra SS, Nasrallah FA. The detection of mild traumatic brain injury in paediatrics using artificial neural networks. Comput Biol Med. 2021 Aug;135:104614.
22. Liu LY, Bush WS, Koyutürk M, Karakurt G. Interplay between traumatic brain injury and intimate partner violence: data driven analysis utilizing electronic health records. BMC Women's Health. 2020 Dec 7;20(1):269.
23. Hossain E, Rana R, Higgins N, Soar J, Barua PD, Pisani AR, et al. Natural language processing in electronic health records in relation to healthcare decision-making: a systematic review. Comput Biol Med. 2023 Mar 1;155:106649.
24. Bird S, Klein E, Loper E. Natural language processing with Python: analyzing text with the Natural Language Toolkit. O'Reilly Media, Inc.; 2009. 506 p.
25. Jacobson O, Dalianis H. Applying deep learning on electronic health records in Swedish to predict healthcare-associated infections. In: Proceedings of the 15th Workshop on Biomedical Natural Language Processing [Internet]. Berlin, Germany: Association for Computational Linguistics; 2016 [cited 2023 Aug 14]. p. 191–5. Available from: https://aclanthology.org/W16-2926
26. Mouselimis L. fuzzywuzzyR: fuzzy string matching. [Internet]. 2021. Available from: https://CRAN.R-project.org/package=fuzzywuzzyR
27. SeatGeek Inc. seatgeek/fuzzywuzzy [Internet]. 2023 [cited 2023 Aug 14]. Available from: https://github.com/seatgeek/fuzzywuzzy
28. Breiman L. Random forests. Mach Learn. 2001 Oct;45(1):5-32.
29. Mikolov T, Grave E, Bojanowski P, Puhrsch C, Joulin A. Advances in pre-training distributed word representations. arXiv preprint arXiv:1712.09405. 2017 Dec 26.
30. Sanyal J, Tariq A, Kurian AW, Rubin D, Banerjee I. Weakly supervised temporal model for prediction of breast cancer distant recurrence. Sci Rep. 2021 May 4;11(1):9461.