

Paper Review

Shauna Heron
Laurentian University

Introduction and Motivation

This review analyzes a study by R. Garriga, T. S. Buda, J. Guerreiro, J. O. Iglesias, I. E. Aguerri, and A. Matic [1], published in *Cell Medical*, titled “*Combining Clinical Notes with Structured Electronic Health Records Enhances the Prediction of Mental Health Crises.*”

The research investigates the utility of combining unstructured clinical notes with structured data from electronic health records (EHR) to improve the prediction of mental health crises [1]. The relevance of this research is underscored by a global rise in mental health-related hospitalizations coinciding with significant workforce challenges, stressing the need for tools that might anticipate demand and help manage caseloads more efficiently. Even in Ontario, recent data highlights that hospitalizations for mental health conditions, particularly among youth, have surged since the COVID-19 pandemic [2]. Moreover, those requiring hospitalization (e.g., emotional breakdowns, substance overdoses and suicide attempts) have increased by 90% [2]—many of which many might have been prevented with early intervention, suggesting that an automated tool for predicting crisis onset might help anticipate and address mental health crises before they peak.

As the authors point out, leveraging EHRs to bolster clinical decision-making is not new [1]. Clinicians and researchers have long utilized structured data like diagnosis codes, lab results, and medication records to inform predictive models. However, computational limits as well as a lack of sophisticated modeling tools necessary to implement such models meant that unstructured data like clinical notes and other free-form text, were most often left out of feature sets. Considering the rich contextual information that clinical notes contain, as well as advances in free-form text processing, Garriga et al., set out to test whether including clinical notes alongside structured EHR data would enhance predictive performance of mental health crisis prediction compared to structured data alone.

Hybrid Solution

With this gap in mind, the authors proposed a hybrid approach that built on a previous study [3] that used structured EHR to predict a weekly crisis-risk score. This time, they included unstructured data in the form of clinical notes to bolster the predictions. Specifically, a deep neural network (DNN) was implemented to integrate both data types, hypothesizing that models including unstructured data (such as clinical notes) would predict the probability of a crisis event more accurately than just structured data alone.

The models they compared included a deep neural network trained on structured data (referred to as Struct DNN), a model trained exclusively on unstructured data (Text DNN), and a combined model that integrated both structured and unstructured data (Hybrid DNN).

Critically, the results of the earlier tests were used to build a final ensemble model that utilized predictions from the Hybrid DNN when unstructured data was available and predictions from the Struct DNN model otherwise (referred to as Ensemble DNN). See Figure 1 below for a flow chart of the experimental and modeling process.

Model Architecture and Data Flow Diagram:

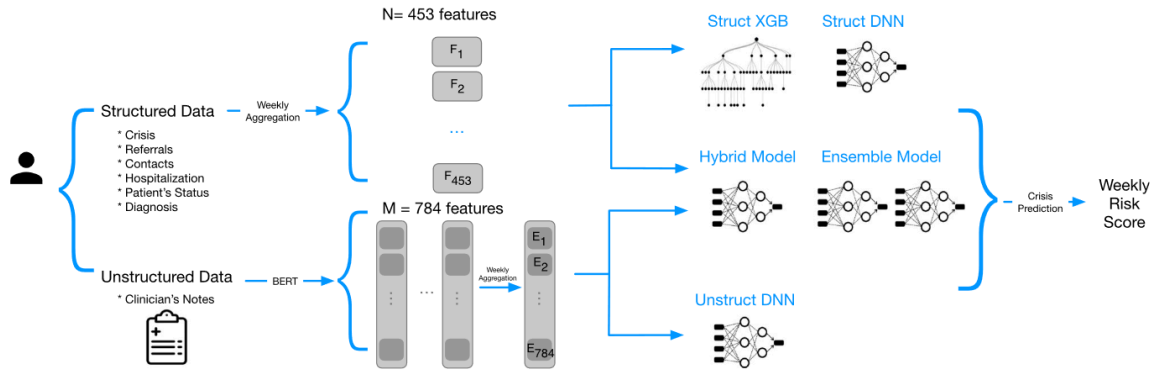


Figure 1. Overview of the Hybrid DNN model and data input structure, showing how structured and unstructured data are processed. Adapted from [1, pg. 3]

To address the issue of class imbalance (since relapses were rare with only 1.3% prevalence), models were tuned to maximize the area under the precision-recall curve (AUPRC)—which as we learned in class is particularly useful metric when evaluating results produced by imbalanced classes. As highlighted in our lessons, precision-recall metrics provide a better measure of a model’s predictive quality than accuracy when data is unbalanced and in settings where sensitivity and an ability to minimize both false positives and false negatives are critical to efficiently predicting health crises where mistakes can be costly.

To evaluate model performance, two baselines were used for comparison: a 5-factor logistic regression model informed by significant variables suggested in prior literature (LogReg5) and a heuristic model ranking patients by the total number of crises experienced in the past year.

Experimental Results

The experiments revealed that the best model trained on structured data was an XGBoost classifier, a tree-based model implementing gradient boosting. However, for models that included unstructured data (clinical notes), the Hybrid DNN provided the best performance. This finding suggests that their hypothesis was correct: incorporating clinical notes improved the model’s ability to identify potential mental health crises earlier compared to models trained on structured data alone. The Hybrid DNN also outperformed both the logistic regression baseline and the heuristic model with the highest AUPRC.

Performance Comparison Across Models:

The comparative performance of the Structured-XGBoost, unstructured DNN and the hybrid DNN that combined both data types, is illustrated by the AUROC and AUPRC values for model performance at 10-week intervals across the 52 weeks.

When looking at the AUPRC (area under the precision-recall curves) below, note the increase in performance across models at each 10-week interval. Note also how small the values are—this is common in health data when predicting events that are rare (like crisis events) [4].

According to the author's the increase over time is in part due to the fact that information in client record increases over time, however the increase may be more related the clinical fact that the prevalence of crisis-relapse tends to increase over time regardless. Nevertheless, note that the structured XGBoost model outperforms both models at 10-weeks, but at 20 weeks the overall performance of the XGBoost model is surpassed by the hybrid model. [say something here about the precision recall values]

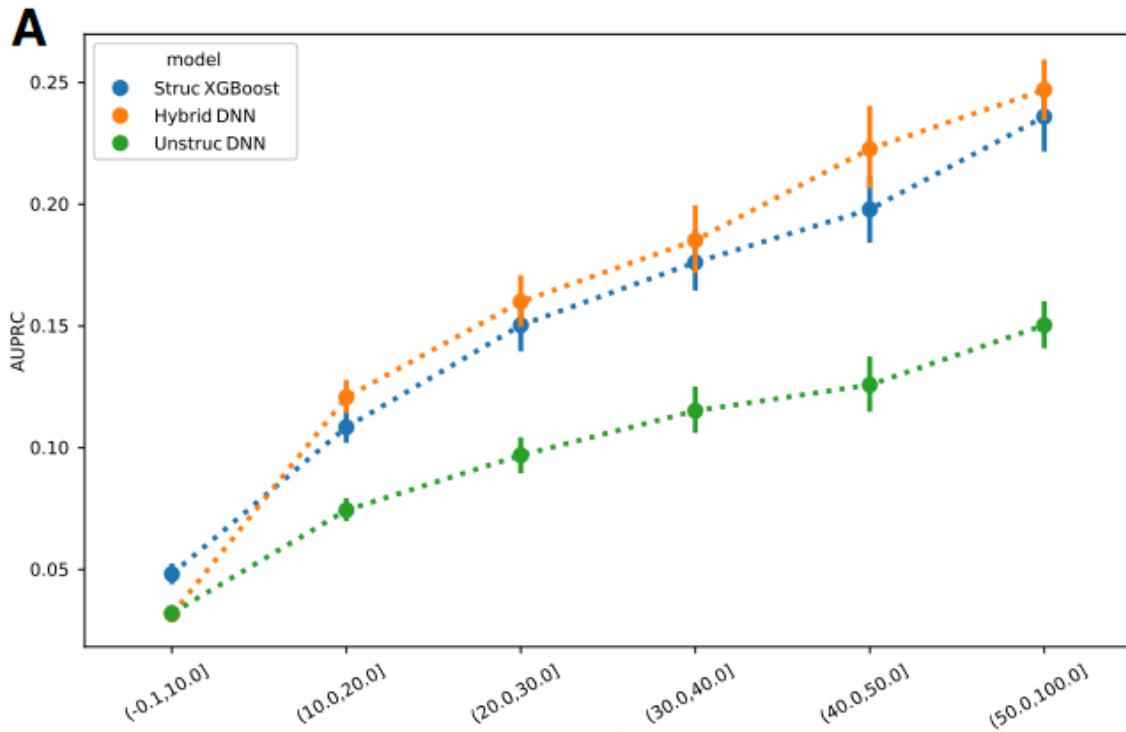


Figure 1: Figure: Precision-Recall Curves

Figure 3. Precision-recall curves for the Hybrid DNN, Struct DNN, and Text DNN, across 10-week intervals illustrating an increase in performance of all models over time as the amount of information increases across weeks. Adapted from [1, pg. 5];

The AUROC curve on the other hand shows a different pattern. The performance of the structured XGBoost and unstructured DNN models both *decrease* over time until 40 weeks when both the structured and hybrid model increase in performance. Interestingly, the performance of the un-

structured-only model decreases across the entire 52 weeks, which the authors suggest highlights the complexity of modelling mental health crises: while the quantity of data in a given client file tends to increase over time, the overall number of patients requiring treatment this long *decreases* which impacts performance of the neural network (i.e., text input tends to require more examples or training instances to perform adequately in NNs). Note too how the structured-only xgboost is best at predicting crisis with fewer than 10% of notes, but with 20% of weeks with available notes, the hybrid model's performance offers a small improvement on performance. This indicates that the addition of unstructured data helped the model make more accurate predictions across a range of thresholds, improving its overall discriminatory power. This chart highlights the benefits of implementing the final ensemble model combining the structured and hybrid models depending on the availability of data in the EHR. The ensemble approach provides flexible capability utilizing the data at hand in earliest stages but with more data can update its routine utilizing both types of data.

AUROC Across 52 weeks:

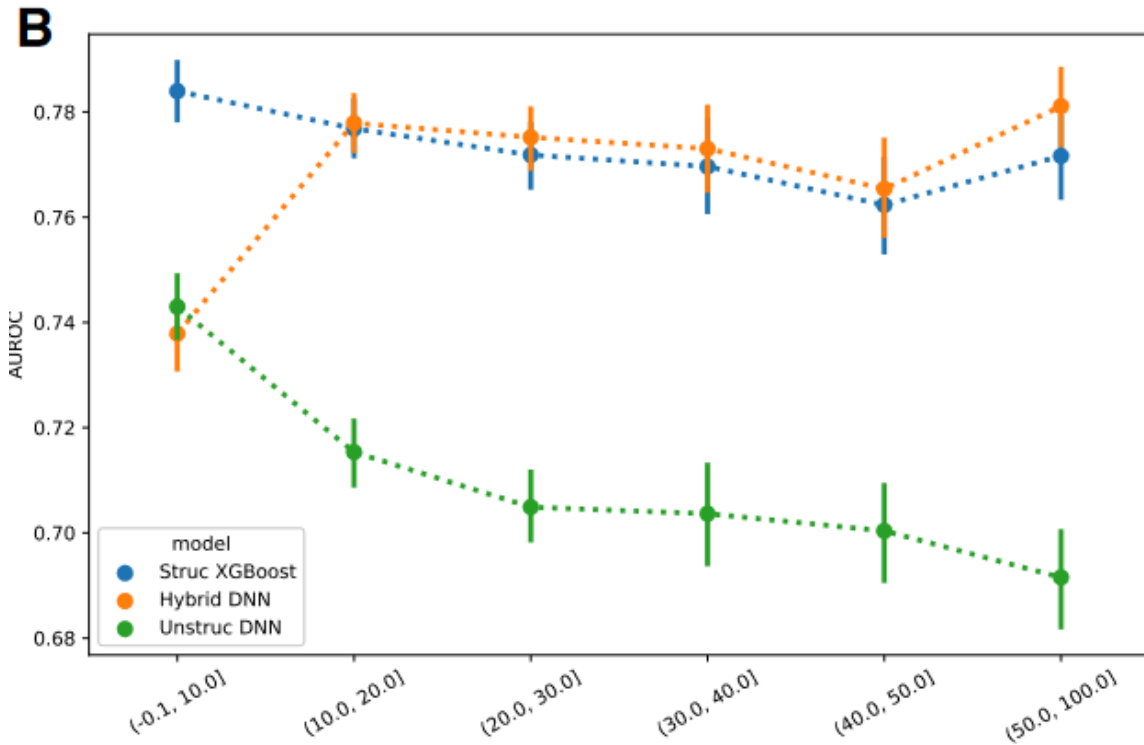


Figure 4. Performance of the Hybrid model compared to the unstructured and structured only models in predicting crisis episodes at 10 week intervals. Points and lines indicate mean and \pm standard deviation values computed in the 52 weeks of the test set. Adapted from [1, pg. 5];

Critical Analysis

What did the paper do well?

- excellent report; reproducible, resource sharing, code sharing, SHAP analysis is rare to see love how that was included in this paper.
- Difficult to find areas that were weak.FIn

The paper offered a compelling case for the inclusion of unstructured data to bolster structured data in predictive modeling of healthcare data, highlight several important techniques including the use of an ensemble of models depending on the data available as well as demonstrating the importance of feature analysis which allows both cohort-level and instance-specific explanations which is important for enhancing transparency in healthcare machine learning models.

an ensemble approach and demonstrated al. Firstly, while the authors highlighted the potential benefits of unstructured data, the interpretability of their hybrid model remained limited. In class, we learned about techniques like Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP), which could be used to make the DNN's predictions more interpretable. Incorporating such methods would make the model's predictions more actionable for healthcare practitioners, who need clear explanations to make informed decisions.

Future research could explore the generalizability of these findings to other domains within healthcare. In particular, in my own research I am wondering how this method could be used to predict case complexity to caseload management. and refine R. Garriga, T. S. Buda, J. Guerreiro, J. O. Iglesias, I. E. Aguerri, and A. Matic [1] 's approach to make the integration of unstructured data more interpretable. Another potential direction was improving methods for natural language processing (NLP) to better extract meaningful insights from clinical notes, which could further enhance predictive performance.

I especially liked their inclusion of SHAP values and analyses of variables that contributed most to model predictions–TALK MORE ABOUT THIS.

Secondly, the study primarily focused on a single health system, which might limit the generalizability of the results. Future studies should consider multi-site evaluations to determine if the predictive power of these models held in different settings with varied data quality and clinical practices.

Lastly, while the results were promising, they relied heavily on high-quality, well-documented clinical notes. In practice, the quality of clinical notes could vary significantly, and the models' reliance on unstructured data could lead to inconsistencies in predictions if the notes were incomplete or ambiguous.

Conclusion

The study effectively demonstrated that incorporating unstructured data, such as clinical notes, alongside structured EHR data improved the ability to predict mental health crises. Importantly their findings underscore the potential for including qualitative clinical information in predictive modeling and highlights a novel ensemble methodology that can be used to build a flexible, combined model that can extract value from both data types. depening on what is available in the client record.

Bibliography

- [1] R. Garriga, T. S. Buda, J. Guerreiro, J. O. Iglesias, I. E. Aguerri, and A. Matic, “Combining clinical notes with structured electronic health records enhances the prediction of mental health crises,” *Cell Reports Medicine*, vol. 4, no. 11, Nov. 2023, doi: 10.1016/j.xcrm.2023.101260.
- [2] CMHO, “Addressing Urgent Workforce Challenges in Child and Youth Mental Health,” Mar. 2022.
- [3] R. Garriga *et al.*, “Machine learning model to predict mental health crises from electronic health records,” *Nature Medicine*, vol. 28, no. 6, pp. 1240–1248, Jun. 2022, doi: 10.1038/s41591-022-01811-5.
- [4] “Measuring Performance: AUPRC and Average Precision.” [Online]. Available: <https://glassboxmedicine.com/2019/03/02/measuring-performance-auprc/>