GR-63

Adaptive Attention Aware Fusion for Human-in-Loop Behavioral Health Detection

Martin Brown, Md Abdullah Al Hafiz Khan, Dominic Thomas, Monica Nandan, Yong Pei

KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING

Abstract

Identifying behavioral health is paramount for law enforcement officers to provide appropriate follow-up community care. Currently, officers manually identify behavioral health cases for follow-up designation. We developed an AI tool to automatically detect these cases from police narrative reports. Our model leverages contextual and semantic information from the reports and relevant behavioral health keywords cues as keywords. Additionally, we present a novel human-in-the-loop uncertainty-based querying strategy that selects the most informative and diverse samples for expert annotation to actively teach our proposed model.

Introduction

911 crises involving individuals with behavioral health issues have a high likelihood to reoccur and worsen in current systems [1-4]. Many police forces are implementing co-response teams that aim to offer alternative interventions rather than allowing situations to worsen, resulting in increased emergency calls [5,6]. Police require tools to screen through high volumes of these crisis cases to suggest instances that might benefit from alternate, improved responses to avoid worsening and further crises [7,8]. Therefore, we develop and present a data analytic solution for this problem, tested on actual data from a partner police force in a large, suburban context in the Southeastern United States. Figure 1 presents where the behavioral health detection tool fits in the crisis workflow.

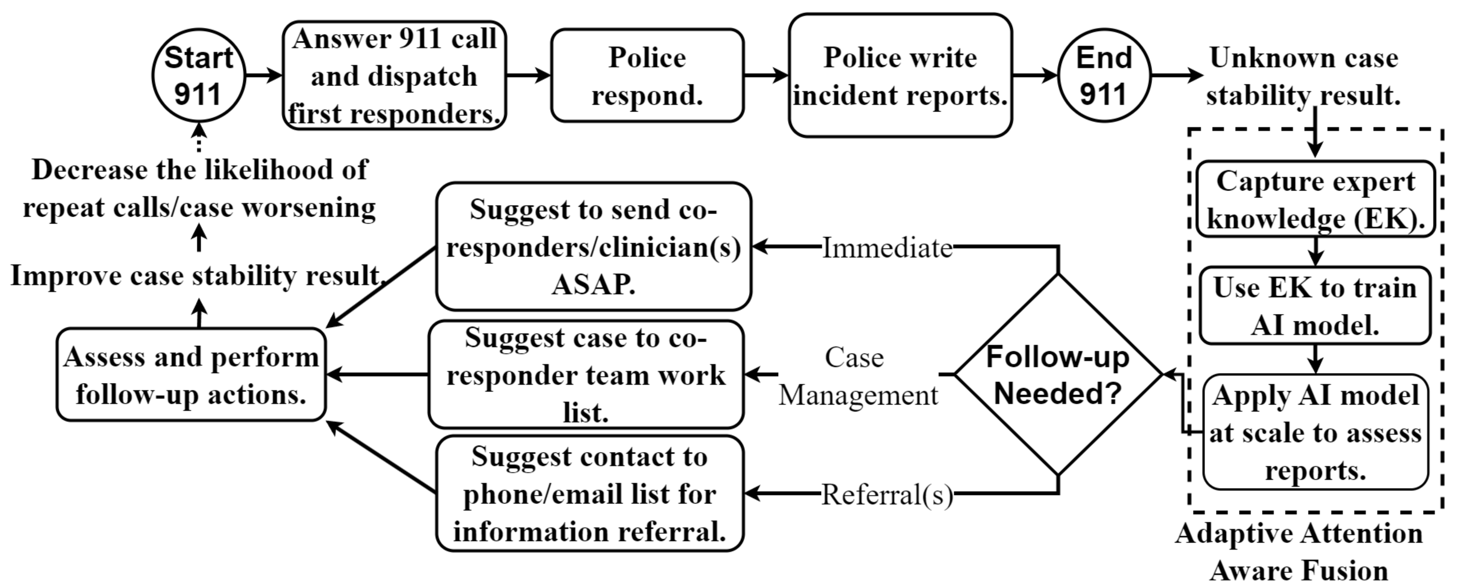


Fig. 1. Project workflow and tool integration (right). This project automates post 911 incidents to augment the existing process for faster, more equitable assessments.

Identifying behavioral health issues are challenging as they require careful diagnosis and regulation [8]. Behavioral health signals within 911 crisis data exist in textual form outside of formal standards [9]. Consequently, the data tends to be noisy and unstructured, presenting faint behavioral health signals. An AI solution would need to be transparent and fair to mitigate discriminatory tendencies. We present a novel model incorporating AI with NLP and ML, designed to address the challenges with three major contributions.

- We initially used a limited set of ground truth samples, resulting in low performance, but improved results significantly by employing human-in-the-loop active learning with an innovative querying strategy known as **uncertainty-based informative cluster querying**.
- The **adaptive attention aware fusion**, fuses a typical context attention-based approach with behavioral health-specific keyword cues to ensure explainability, interpretability, and boost accuracy.
- We configured the model to disregard a list of demographic words to ensure no adverse discriminatory stereotyping results.

Police Report Data Document (D)

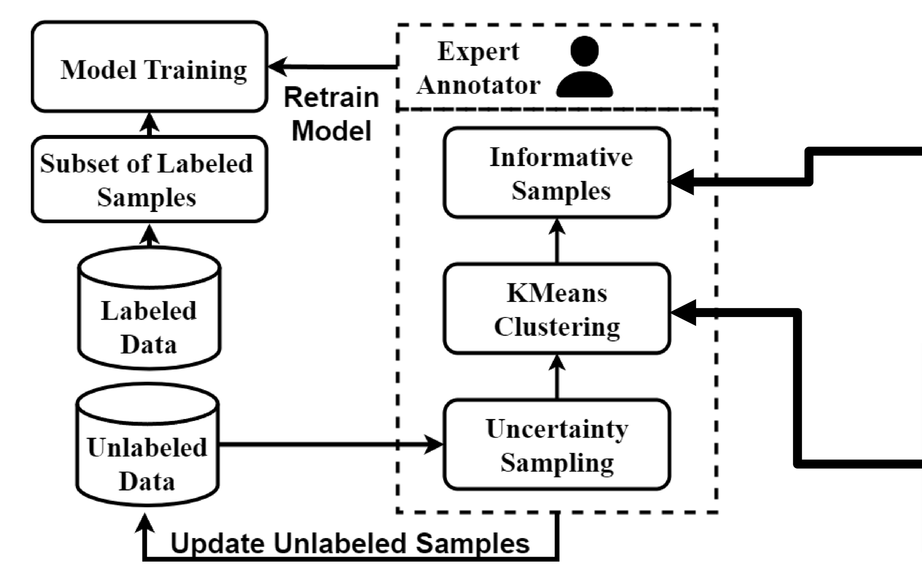
Behavioral Health Sub-Category	Example
Mental Health	Involving an individual with a diagnosed mental disorder, like schizophrenia or suicidal ideations.
Domestic/Social	Involving multiple individuals in a home setting, like husband/wife domestic disputes.
Non-Domestic/Social	Involving multiple individuals not in a home setting, like committing crimes on unrelated perpetrators.
Substance Abuse	Individuals with persistent drug/alcohol abuse problems.

Informative Cluster Querying for Active Learning

Uncertainty: To reduce the number of samples for clustering and querying we apply uncertainty model predictions.

Clustering: Elkan's algorithm is used to enhance clustering performance within k-means.

Truncation: For NLP, clustering is performed on a truncated portion of the text to reduce dimensionality.



Informative using cluster entropy:

$$H(U) = argmax_U \left(-\sum_C P(C|U) \log P(C|U) \right)$$

$$f_C(U) = argmax_U \, (H(U) \cdot S_C(U))$$

Top n samples are queried from this function

Outlier are determined through Silhouette score:

$$S_C(U) = {a - b}/_{\max(a, b)}$$

'a' is the mean intra-cluster distance.
'b' is the mean nearest cluster distance.

Adaptive Attention Aware Fusion Architecture

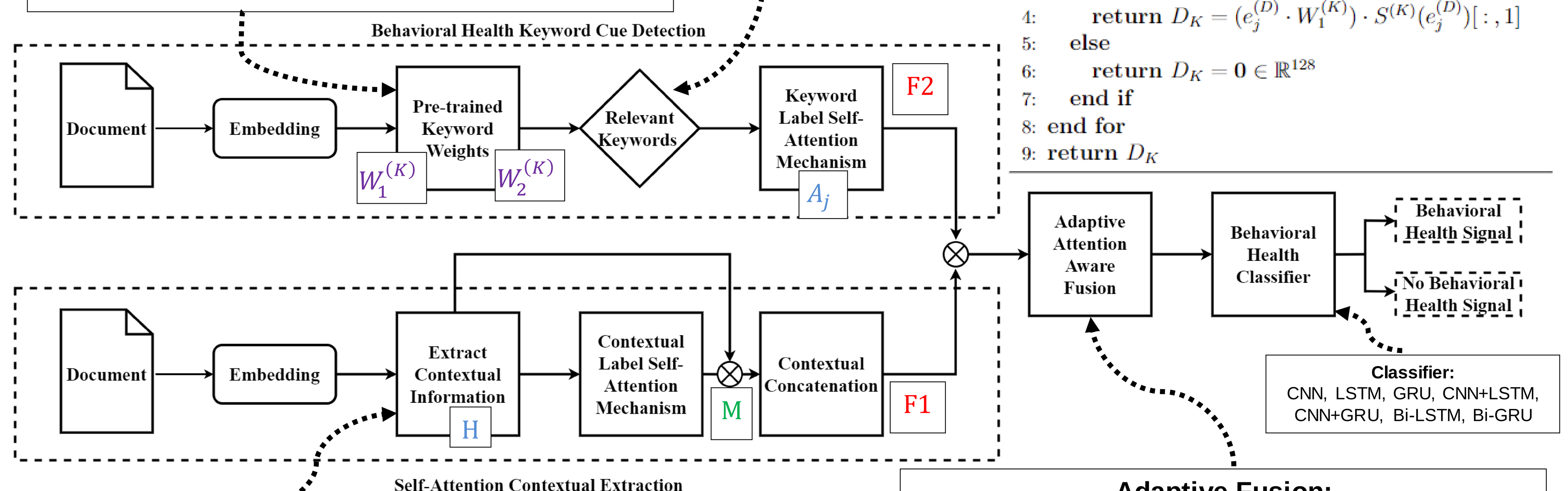
Pre-train model to related keywords (K) to their categories:

$$S^{(K)}(kw_i) = Softmax\left(ReLU\left(e_i^{(K)} \cdot W_1^{(K)}\right) \cdot W_2^{(K)}\right).$$

$W_1^{(K)}$ and $W_2^{(K)}$ fixed in our model

Algorithm 1 'Relevant Keyword' Decision Algorithm

Require: $W_1^{(K)}, W_2^{(K)}$
1: **for** j $= 1, 2, \ldots, M = 500$ **do**
2: **Generate:** $S^{(K)}(e_j^{(D)})$
3: **if** $S^{(K)}(e_j^{(D)})$ is not 'None' **then**
4: **return** $D_K = (e_j^{(D)} \cdot W_1^{(K)}) \cdot S^{(K)}(e_j^{(D)})[:, 1]$
5: **else**
6: **return** $D_K = 0 \in \mathbb{R}^{128}$
7: **end if**
8: **end for**
9: **return** D_K



Bi-LSTM layer to extract contextual information:

$$\vec{h}_t^{(D)} = LSTM^{(D)}\left(e_j^{(D)}, \vec{h}_{t-1}^{(D)}\right). \ \overleftarrow{h}_t^{(D)} = LSTM^{(D)}\left(e_j^{(D)}, \overleftarrow{h}_{t-1}^{(D)}\right).$$

$$\vec{H}^{(D)} = \left\{\vec{h}_t^{(D)}\right\}_{t=1}^{dim}. \ \overleftarrow{H}^{(D)} = \left\{\overleftarrow{h}_t^{(D)}\right\}_{t=1}^{dim}. \ H^{(D)} = \left(\vec{H}^{(D)}, \overleftarrow{H}^{(D)}\right).$$

Adaptive Fusion:

$$F1 = \sum_{j=1}^{100} X_j^{(1)} \cdot M + b^{(1)}. \ F2 = \sum_{j=1}^{100} X_j^{(2)} \cdot A_j^{(D_k)} + b^{(2)}.$$

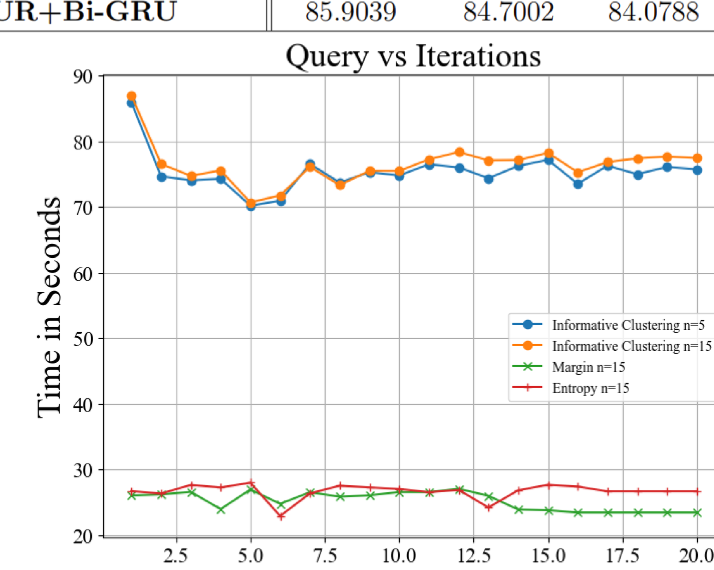
$$\alpha = Sigmoid(F1, X^{(3)}). \ \beta = Sigmoid(F2, X^{(4)}).$$

constrained $\alpha + \beta = 1$

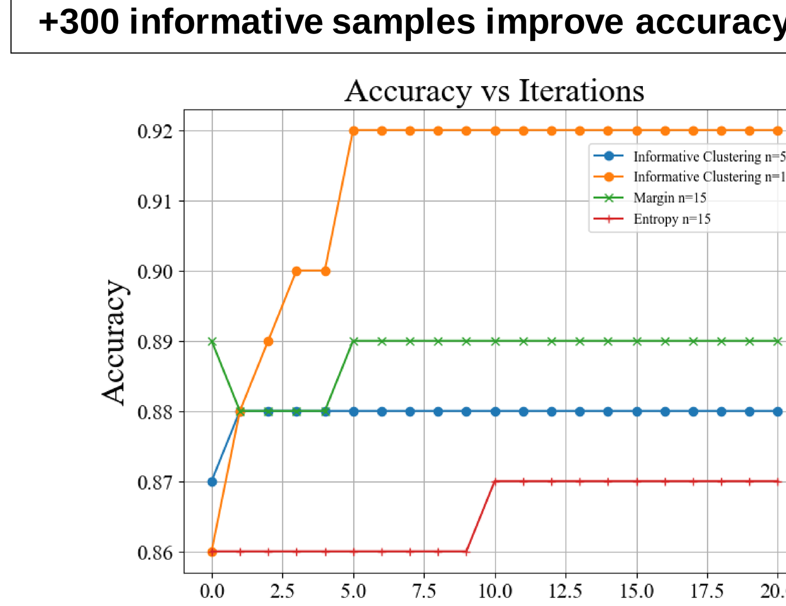
Experiment Results

298 ground truth with 33% Behavioral Health Related

Classifier	Accuracy	Precision	Recall	F1
EMB+CNN	86.2316	86.8694	82.1154	83.6882
OUR+CNN	**87.5763**	**87.3077**	**85.0806**	**85.6667**
EMB+LSTM	69.1299	70.8505	54.7024	49.1326
OUR+LSTM	86.2429	85.4872	83.8178	84.3229
EMB+GRU	69.1243	74.0917	54.4524	48.5217
OUR+GRU	86.9040	86.0565	85.0549	85.1877
EMB+CNN+GRU	69.4633	74.2183	54.9524	49.3337
OUR+CNN+GRU	84.5254	84.4639	81.0137	80.5264
EMB+CNN+LSTM	71.8023	79.3836	58.8269	56.3297
OUR+CNN+LSTM	76.8474	72.5957	68.3919	65.8983
EMB+Bi-LSTM	86.5819	85.5291	84.8590	84.8783
OUR+Bi-LSTM	86.5989	85.6504	83.8416	84.5643
EMB+Bi-GRU	85.2260	84.7431	82.7683	83.0886
OUR+Bi-GRU	85.9039	84.7002	84.0788	84.1722



+300 informative samples improve accuracy to 92%



Highlighted from weights: $W_1^{(K)}$ and $W_2^{(K)}$

'On DATE1 I responded to FAC1 reference a child custody dispute. Upon arrival I spoke with PERSON1 who informed me she was trying to get custody of her children from her husband PERSON2 who would not release them. PERSON1 then provided me with civil paperwork which showed she was supposed to have custody of the children at that time. I informed PERSON1 this was a civil matter and I could not make PERSON2 release the children. PERSON1 advised she understood. I then spoke with PERSON2 who advised he was not going to give the children to PERSON1 because it was father`s day. I informed PERSON2 if he did not release the children he could be found in violation of a civil order to which PERSON4 advised he was not releasing the children.'

Behavioral Health Keywords (K)

Keyword Category	Description	Examples
Situation (SIT)	Words to explain a situation, including violence tendencies	abuse, aggressive, dementia, mental
Child (CHI)	Words related to juvenile	child, youth, toddler, kid
Crime (CRI)	Words related to criminal acts	robbery, shoplift, kidnap, battery
Disposition (DIS)	Words for what happened before/after 911	arrest, morgue, hospital, referral
Drugs (DRU)	Words related to drug involvement	cocaine, drug, dope, meth
Medication (MED)	Words related to prescription medication	adderall, antidepressant, xanax, zoloft
Demographic (DEM)	Words related to demographic information	black, white, asian, man, woman

Conclusions and Limitations

Conclusions: Our adaptive attention-aware fusion surpasses existing classifiers, achieving 87.58% accuracy and 85.67% F1-score on 300 annotated reports. With our novel querying strategy, accuracy increases to 92%, F1-score to 91.1%. On unseen samples, accuracy remains high at 93.75%, F1-score at 93.61%. Notably, our model offers interpretability by extracting and highlighting associated keywords for each behavioral health category. Lastly, our proposed model effectively emphasizes keywords within documents crucial for behavioral health detection through the internal trained weights.

Limitations: Effective training of the model within the active learning framework initially necessitates a larger pool of initial samples to ensure a diverse querying process, enhancing model annotation and performance. Iterative active learning is crucial to fully leverage the adaptive attention-aware fusion model with newly labeled samples, ensuring continuous improvement. Employing a grid search method to determine the optimal number of samples to query per iteration is essential for maximizing training efficiency, although scaling up the model's training requires additional computational resources to meet increased data demands.

Acknowledgments

We would like to express our special thanks to external Co-Responder Program partners Police Officer Jacob King from Cobb Path Team and Mr. Matthew Dames, a Community Response Behavioral Health Clinician, who provided valuable feedback throughout the behavioral health detection. Their collaborative involvement significantly contributed to the success of our research. Moreover, we wish to extend our appreciation to the Institutional Review Board (IRB) at Kennesaw State University, represented by the code IRB-FY21-227, for ensuring and upholding the ethical integrity and compliance standards of our study. Lastly, our sincere appreciation goes to the annotators and subject matter experts whose unwavering support significantly contributed to the successful completion of our research.

Contact Information

Martin Brown: mbrow612@students.Kennesaw.edu
Md Abdullah Al Hafiz Khan: mkhan74@kennesaw.edu
Dominic Thomas: dthom310@kennesaw.edu
Monica Nandan: mnandan@kennesaw.edu
Yong Pei: ypei@kennesaw.edu

References

[1] Castner, J., Wu, YW., Mehrok, N., Gadre, A., Hewner, S. (2015). Frequent emergency department utilization and behavioral health diagnoses. *Nurse Research*, 64(1), 3-12. https://doi.org/10.1097/nnr.0000000000000065
[2] Substance Abuse and Mental Health Services Administration. (n.d). SAMHSA - Behavioral health integration. https://www.samhsa.gov/sites/default/files/samhsa-behavioral-health-integration.pdf
[3] National Institute of Mental Health. (2023). Mental Illness. https://www.nimh.nih.gov/health/statistics/mental-illness
[4] Vera Institute of Justice. (2022). *911 Analysis: How Civilian Crisis Responders Can Divert Behavioral Health Calls from Police*. https://www.vera.org/downloads/publications/
[5] Lawrenceville Police Georgia. (n.d.). *Co-responder program*. https://www.lawrencevillega.org/624/Co-Responder-Program
[6] California Legislative Information \& Georgia General Assembly. (2023, September 7). *Senate Bill 403*. https://www.legis.ga.gov/api/legislation/document/20212022/211478
[7] Wilson, J. M. (2022). *Reframing the police staffing challenge: Strategies for meeting workload demand. Community Oriented Policing Services,* 15(1). https://cops.usdoj.gov/html/dispatch/01-2022/police_staffing_strategies.html}
[8] American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th edition). https://doi.org/10.1176/appi.books.978089042559
[9] Hashmi, T., Thomas, D., \& Nandan, M. (2023). First responders, mental health, dispatch coding, COVID-19: Crisis within a crisis. *Journal of Emergency Management*, 21(3). https://wmpllc.org/ojs/index.php/jem/issue/view/379
[10] Krishnan, J. (2023, June 9). The role of AI in mental health. *Active Minds Blog*. https://www.activeminds.org/blog/exploring-the-pros-and-cons-of-ai-in-mental-health-care/
[11] Cost, B. (2023, March 30). *Married father commits suicide after encouragement by AI chatbot: Widow*. New York Post. https://nypost.com/2023/03/30/married-father-commits-suicide-after-encouragement-by-ai-chatbot-widow/
[12] Ji, S., Zhang, T., Ansari, L., Fu, J., Tiwari, P., and Cambria, E. (2021). MentalBERT: Publicly available pretrained language models for mental healthcare. *arXiv*. https://doi.org/10.48550/arXiv.2110.15621
[13] Brown M., Khan, M. A. A. H., Thomas, D., Pei, Y., \& Nandan, M. (2023). Detection of behavioral health cases from sensitive police report narratives. In *2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*, 1398--1403. https://doi.org/10.1109/COMPSAC57700.2023.00213