A Trustworthy Classifiers Chain for Cyberbullying Detection Using Bystander Roles

1st Haifa Saleh Alfurayj
School of Computer Sciences
Universiti Sains Malaysia
Penang, Malaysia
School of Computer Sciences
Qassim University
Qassim, Saudi Arabia
0000-0002-6875-5755

2nd Syaheerah Lebai Lutfi
Dept of Medical Education & Medical Informatics
Sultan Qaboos University
Sultanate of Oman
School of Computer Sciences
Universiti Sains Malaysia
Penang, Malaysia
0000-0001-7349-0061

***Abstract*—Cyberbullying has become a pressing societal challenge in the era of digital communication. Cyberbullying language is often part of a longer conversation that involves bystander roles. While incorporating additional contextual information—such as bystander replies—offers a more realistic and effective strategy for identifying instances of cyberbullying, most current studies are limited by their reliance on interpreting standalone Tweets. In this work, we explore fine-grained cyberbullying detection using bystander information through a chained model. However, sequential classification in chained models is prone to error propagation, where errors in earlier layers negatively affect the training and predictions in subsequent layers. Studies primarily focus on optimizing the chain classifier by developing strategies for selecting the optimal labels ordering. To improve the chained model with the pre-specified order as in our approach, it is crucial to enhance the reliability of the earlier layers to mitigate error propagation. This paper addresses this issue by incorporating a filter layer that determines the optimal threshold based on confidence and permissible error. This approach ensures that only predictions above the specified threshold are considered *confident* and carried over for use in the next layer of the chain. The results demonstrate that, using the *Confident subset* to train the upper classifier in the chain offers a distinct advantage over the standard chain model.**

***Index Terms*—cyberbullying, bystander, chain, confidence, detection**

I. INTRODUCTION

Cyberbullying involves real-world situations with multiple events initiated by a group. In social network sites, the events interpreted inform of main post by the first author and the replies by the bystander. Bystander roles can be categorized as Instigators, Impartials, Defenders, and Others as defined in [1]. To address the risks associated with cyberbullying, increasing attention has been directed toward the development of automated detection models. While these efforts have led to the creation of numerous models, most of them are limited by their reliance on interpreting standalone Tweets. This approach can be particularly challenging when the Tweet is part of a longer conversation. Incorporating additional contextual information—such as bystander replies within ongoing conversations—offers a more realistic and effective strategy for identifying instances of cyberbullying.

Identifying the roles of bystander in conversation threads is crucial for enhancing the fine-grained cyberbulying detection. Research suggests using multi-labels learning approaches, which explicitly consider label interdependence, generally results in better predictive performance [2], [3].

Cyber-aggression is defined as aggressive behavior intended to cause harm to a person (e.g., name-calling by an anonymous online user). Cyberbullying is defined as frequent aggressive behavior carried out electronically by a person or a group of people, aimed at inflicting harm on a person who cannot easily fight back, creating a power imbalance in which the bully has power over the victim [4]–[7]. Using the above definitions, we explore frequency and power imbalance are distinguishing characteristics that differentiate cyberbullying instances from cyber-aggression. Identifying bystander roles can help distinguish between cyber-aggression and cyberbullying. Bystander behaviors often reveal severity, frequency, and power dynamics of the interaction. For example, the presence of instigating bystanders may indicate frequent harm, and may also signal a power imbalance where the victim is unsupported.

Multi-label Classification (MLC) is a type of supervised learning problem where each instance can be linked to multi-labels. MLC captures the attention of Machine Learning researchers due to its applicability to a wide variety of applications, such as text classification [8], emotion recognition [9], bank marketing [10], acoustic event detection [11], and multi-disease risk prediction [12]. The difference between MLC and Single-label Classification (SLC) is that MLC predicts multiple labels, whereas the conventional task of SLC involves predicting just one class label. The SLC and MLC can be either binary or multiclass, depending on the number of classes involved in the classification task.

MLC problems can be solved by one of two approaches: algorithm adaptation or problem transformation. Algorithm adaptation is the most straightforward method that modifies SLC methods to be suitable for multi-label problems. The problem transformation method, on the other hand, transforms the multi-label problem into one or more single-label classification problems [2]. Problem transformation method

has two popular ways - Classifiers Chain (CC) or Binary Relevance. Binary Relevance is the more common approach due to its simplicity; it independently trains a binary classifier for each label, overlooking the explicit interaction among events. On the other hand, the CC method employs multiple SLC classifiers equal to the number of labels, with each trained for a specific label. To perform classification for a new instance, CC begins by predicting the value of the first label. Then, it takes this instance together with the predicted value as the input to predict the value of the next label. This process continues until the final label is predicted [3]. The CC model is widely adopted and popular for its ability to address label dependency, simplicity, and promising experimental results.

CC performance suffers from error propagation problems [13], meaning that errors generated by earlier classifiers can propagate to subsequent classifiers in the chain, leading to additional errors. To reduce the influence of error propagation issues, this research proposes a Trustworthy Classifiers Chain by adding the proposed filter layer. This layer optimizes model performance and limits error propagation by managing both confidence and error.

In Section II, we provide an overview of the existing studies enhancing the Classifiers Chain model. The proposed Classifiers Chain model is introduced in Section III-B. In Section III-C, we present the implementation of the system. The results of our work are discussed in Section IV, while Section V provides directions for future work and concludes the paper.

II. BACKGROUND AND RELATED WORK

Machine Learning researchers are actively working to mitigate the limitations of CC while preserving its high performance for complex multi-label problems. CC performance is affected by label ordering because different Classifiers Chains involve different numbers of various classifiers trained on different training sets [2], [3], [14].

The enhancement of Classifiers Chains (CC) was initiated by [3], who introduced an ensemble Classifiers Chain to average the multi-label predictions of CC over a set of random chain orderings in order to obtain the optimal label ordering and limit the issue of error propagation. However, the sequence of labels still suffers from strong randomness, which remains a challenge among researchers. As a result, there has been a growing body of work focused on implementing and improving the multi-label Classifiers Chain method by more effectively modeling label correlations, as outlined in Table I. For example, in [15], an label ordering approach based on label dependence measurement strategy was proposed for improving multi-label Classifiers Chain accuracy. The proposed approach is based on mining of the correlation information and association rules from frequent pattern between labels itemsets. In the process of mining association rules, strong rules are identified into account to meet both the minimum support and lift in addition to confidence thresholds. Then a directed acyclic graph is constructed to obtain the learning order of labels arranged to train each classifier. Similarly, [16] utilized a Bayesian network based on conditional entropy to model the dependency of each label on other labels. A modified scoring function, incorporating both the dependency degree and a complexity penalization term, is then used to assess the quality of the Bayesian network and the resulting label order.

To address the problem of random label sequence ordering, which can lead to error propagation, the authors of [17] proposed a hybrid optimization strategy. A genetic algorithm performs a global search over possible label orders, while swarm optimization techniques refine these orders to identify the optimal sequence. Their proposed method slightly improved the predictive performance of the chain classifier against standard CC and Binary Relevance methods. As one of the pioneers of research in this area, [18] aimed to exploit the correlation between all the dataset's labels using Jaccard index for the first iteration and pairwise measure for the rest of iterations. They then applied the ensemble CC model to the ordered subsets of labels. Their approach utilizes deep Classifiers Chains, specifically BERT models, each responsible for predicting the associated set of labels. The issue with ensemble CC related to majority voting is that it overlooks the possibility that some minority learners may produce more accurate outputs, as it does not explicitly address diversity. To overcome these limitations, [19] proposed a chain of SVM learners employing a tournament voting approach, where classifier outcomes compete in groups until one winner remains. They construct training sets using mutual information measures to assess and prioritize data features in relation to the target class variable, retaining only the most significant ones.

In contrast to studies that focus solely on optimizing chaining orders, some research enhances the standard chaining algorithm through different approaches. For instance, the multi-dimensional classification problem is addressed by creating a chain of binary classifiers with one-vs-one (OvO) decomposition [20]. Since the chosen label order influences the performance of CC, this effect is minimized by constructing ensembles of CCs with various orders and combining their predictions using majority voting. Additionally, instead of training all classifiers on the same dataset, it is beneficial to train each CC on a distinct dataset to enhance the diversity of base learners. Furthermore, the authors in [21] incorporated flexibility by using different classifiers within a chain structure. They employed a range of multi-output classification algorithms, such as Random Forest (RF), Decision Trees (DT), and K-Nearest Neighbors (KNN). Their experimental results indicate that their proposed model, using various classifiers, achieved the second-highest overall accuracy among all models and outperformed standard chain-based models in terms of overall accuracy. The authors in [22] handled uncertain labels through two different approaches: the first approach considers all potential scenarios to avoid propagating early uncertain decisions, while the latter approach marginalizes these labels in the predictive model. The study found that as more labels were missing, the accuracy improved.

With the exception of [22], the reviewed studies have

estimated the uncertainty in chain classifiers with the aim of discovering label correlations and determining the optimal order of labels within the chain. The model proposed by [22] computes the overall uncertainty for each label and marginalizes the labels with the highest uncertainty. This method simplifies the prediction process by excluding uncertain labels from direct consideration in the predictive model, thereby avoiding the propagation of uncertainty through the chain and reducing its impact on subsequent labels. However, this approach is not effective for classification problems with a limited number of labels. To improve the chained model with the limited number of labels that pre-specified in order as in our approach, it is crucial to enhance the reliability of the earlier classification stage to mitigate error propagation.

TABLE I
OVERVIEW OF THE STATE-OF-THE-ART CHAIN CLASSIFIERS USED IN VARIOUS FIELDS

Source	Research Problem	Approach
[15]	Optimal labels ordering	Acyclic graph of rules
[22]	Optimal labels ordering	Uncertainty estimation by a convex sets distributions
[16]	Optimal labels ordering	Acyclic Bayesian network based on conditional entropy
[17]	Optimal labels ordering	Particle swarm optimization and a genetic algorithm
[18]	Optimal labels ordering	Labels correlation by Jaccard index and pairwise measure
[19]	Optimal labels ordering	Chain of SVM learners with a tournament voting approach
[20]	Enhances the standard cc	Ensembles of ccs with binary classifiers with one-vs-one (OvO) decomposition
[21]	Enhances the standard CC	Using different classifiers within a chain

III. METHODOLOGY

A. Data Collection

Cyberbullying typically occurs within interactions among multiple individuals, leading to harassment and controversial content. The first person might initiate cyberbullying through a parent post, while bystander escalate it by replying with child posts. Since a standalone post viewed in isolation may not reveal the presence of cyberbullying, it is essential to gather such instances from entire conversation threads that include both the main post and subsequent replies from bystander. Thus, the dataset used was collected from the X social networking site, totaling 13,309 Tweets, consisting of 2,799 threads and 10,510 replies. Users on this platform post Tweets that bystander can reply to, forming a thread.

We collaborated with three experts, each of these experts holds a PhD. They have extensive research experience in various fields, including psychology, organizational behavior, emotional intelligence, and cyberbullying. They understand the difference between cyber-aggression and cyberbullying, but their task was limited to labeling bystander roles and then determining the cyberbullying severity.

The dataset, known as CYBY24, is annotated with two types of labels - the bystander role label and the fine-grained cyberbullying label. The latter is only done for the main post. The labels were defined as follows:

- Fine-grained cyberbullying label with multi-classes corresponding to the 4-point scale (0-1-2-3):
  - Normal
  - Aggression (not bullying)
  - Bullying with low aggression
  - Bullying with high aggression
- bystander roles label with multi-classes:
  - *Instigators* who agree with the thread author of the main post topic.
  - *Defenders* who disagree with the thread topic and who exhibit a defensive manner.
  - *Impartials* who remain neutral or passive.
  - *Others* whose post consent is not related to the thread topic.

As indicated in Table II, a "Normal" discussion thread, users often use online slang words like *"trash"* and *"sucks"* to express their opinions about games and to convey personal views. It can be observed that including bystander replies is important for accurately interpreting the intended meaning, and this feature could also help in better distinguishing between different contexts.

The Fleiss' kappa reliability scores were 0.88 for the bystander role labeling and 0.92 for the fine-grained labeling of cyberbullying, which indicates substantial agreement between experts.

B. fine-grained cyberbullying detection model

In the chained classification model proposed by [1], misclassifications in bystander roles in the first layer might lead to incorrect contextual information being passed to the cyberbullying detection layer, resulting in reduced accuracy and reliability of the overall model. Most earlier studies address this issue by determining the optimal order of the chain. However, in our study we deal with a pre-specified ordering of two labels, the ordering aligns with the data analysis perspective that bystander play an important role in grading the severity of cyberbullying events. So, in this research we propose filter layer to filter the predictions of the first layer to retain only those instances with high confidence.

The proposed fine-grained cyberbullying detection model employs a sequential classification approach that leverages bystander roles information to improve detection accuracy. It comprises three key processes: (1) a transfer learning approach based on a pre-trained BERT model for identifying bystander roles, (2) a proposed filter layer designed to reduce error propagation by retaining only the instances with a confidence score greater than or equal to a specified threshold, and (3) an RNN-based Bidirectional Long Short-Term Memory (BiLSTM) model for classifying fine-grained cyberbullying. The model architecture is illustrated in Figure 1, and the following sections provide a detailed explanation of each component.

TABLE II
CYBY24 DATASET: EXAMPLE OF ANNOTATIONS FOR SOME CATEGORIES

Reply_ID	Tweet Text	Bystander Roles	Cyberbullying Classes
16221	`main tweet:` *The best indicator of racism is if someone likes college basketball over the NBA.*		Normal thread
16221	`reply 1:` *Yup. Check. Important.*	Instigator	
16221	`reply 2:` *NBA is trash compared to college. Just like the NFL.*	Defender	
16221	`reply 3:` *I'm not very knowledgeable about basketball.*	Impartial	
16221	`reply 4:` *NBA has better players (obviously) but overall game structure kinda sucks. Too many 3's, little D, uncalled travels /carries, "hand check" fouls. I tune out until playoffs.*	Defender	

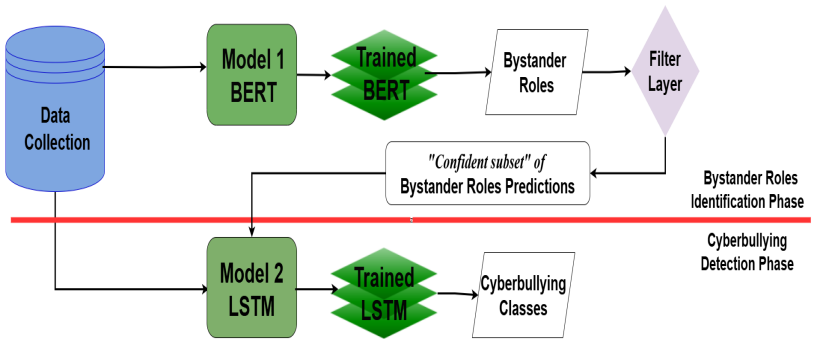

Fig. 1. Flowchart for Chained Cyberbullying Detection Model

C. Model Implementation

process 1: BERT is trained on the input feature (Tweet text) and the bystander roles label to identify bystander roles

process 2: Estimating the uncertainty of predictions made by preceding layer achieved by computing the Maximum Softmax Probability (MSP) from the distribution of logit values. A higher MSP indicates greater confidence in the prediction. Therefore, MSP is often used as a confidence score to determine how certain the model is about a particular prediction.

$$\text{MSP}(x_i) = \max\left(\text{softmax}(z_i)\right) \tag{1}$$

where:

- (x_i) represents the input.
- (z_i) represents the logits.

This approach is introduced in the filter layer to select the predictions that are most likely to be correctly classified. This selection is based on two factors: first, confidence estimation functions, which calculate a confidence measure for each prediction based on the logit values; and second, the calculation of a threshold to filter predictions according to their confidence scores. Based on this threshold, predictions are categorized into two subsets: the *Confident* subset for those predictions that meet the required confidence level, and the *Uncertain* subset for those with lower quality and confidence. The *Confident* subset is used to train the next layer in the chain, while the Uncertain subset is discarded.

The predictions are filtered using specific thresholds. We define a threshold τ as a value that consistently produce the same confident ratio across different datasets. We utilize a function that generates a threshold yielding similar error partitioning. Specifically, a threshold is selected to accepts only E% of the existing errors. The logits are converted into predicted labels by applying the softmax function to generate a probability distribution for each instance, after which the MSP is computed for each instance. The model compares the predicted labels to the true labels to identify instances where the prediction is incorrect. It then creates a list containing the MSP value and the corresponding error status for each instance. This list is sorted by MSP values in descending order. When the number of errors reaches the permissible rate—allowing only E% of the total errors—the model sets the MSP threshold to the MSP value of the current instance. Finally, the predictions are filtered, retaining only those with MSP values above the threshold, resulting in the *Confident* subset.

process 3: BiLSTM is trained on the input features (reply_id, Tweet text), the *Confident* subset of the predicted bystander roles by the first model and the fine-grained cyberbullying label to detect the fine-grained cyberbullying.

The sequential chaining method passes label information between classifiers, allowing the classifier to consider correlations between the labels. A strong correlation between the labels will give the classifier more predictive power. Consequently, the proposed chained model is designed to increase confidence in predictions that will be used for the next phase. More specifically, it aims to filter out classifications that are likely to be misclassified, thereby preventing error

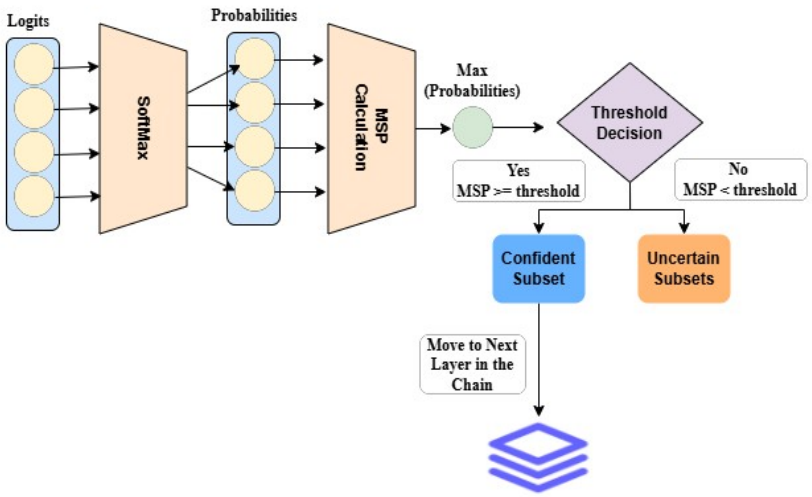

Fig. 2. Zoomed-Out View of the Filter Layer in the Chained Cyberbullying Detection Model

propagation in the sequential classification process.

IV. RESULTS AND DISCUSSION

In this section, we will explain and discuss the results of the evaluation experiments. We conducted experiments to evaluate the detection capability of the model without and with adding the proposed filter layer.

A. Bystander Roles Identification

Table III presents the performance of the first stage of the chained model for bystander role classification, evaluated without applying the filtering layer following BERT fine-tuning.

TABLE III
RESULTS OF PRE-TRAINED LEARNING MODELS EXPERIMENTS ON BYSTANDER ROLES CLASSIFICATION STAGE FOR THE COMPLETE SET OF CYBY24 DATASET WITH #EPOCH = 20.

Model	W-Pc±CI	W-Rc±CI	W-F1±CI
fine-tuning BERT	0.91±0.004	s0.90±0.004	0.90±0.004

Acronyms: W-F1=Weighted-F1, W-Pc=Weighted-Precision, W-Rc=Weighted-Recall, CI=Confidence Interval (95%).

In contrast, Table IV shows the results when the filtering layer is applied to the BERT model. For each permissible error rate (5%, 10%, 15%), it shows the weighted metrics, error rate, and the confident subset size. Compared to the original results without filter layer implication indicated in Table III, there is a significant increase across all evaluated metrics. As proposed, the designed threshold focuses on including instances with a confidence score greater than or equal to the lowest confidence score within an acceptable error margin. This aligns with the proposed method for limiting error propagation by managing confidence and error. This method optimizes the model's performance by determining which predictions are confident and reliable enough to use for training the next layer and which to discard. The size of the confident set increases reasonably with the permissible error rate, which in turn affects the stability of accuracy. As observed, the F1 score remains at 0.98 for error rates of 0.05, 0.10, and 0.15, due to the increase in confident set sizes from 4,235 to 7,587, and then to 9,360, respectively. In all the experiments, the thresholds are almost the same. This highlights the importance of specifying the permissible error rate and then selecting the threshold accordingly, as designed in the proposed method.

TABLE IV
RESULTS OF BERT MODEL EXPERIMENTS WITH FILTER LAYER ON BYSTANDER ROLES CLASSIFICATION STAGE FOR THE COMPLETE SET OF CYBY24 DATASET WITH #EPOCH = 10 IN THREE DIFFERENCE PERMISSIBLE ERROR RATE—ALLOWING ONLY .05 , .10 , .15 OF THE TOTAL ERRORS IN THE CONFIDENT SUBSET.

permissible error	W-F1	Confident Threshold	Error Rate	Confident Subset size	Uncertain Subset size
.05	0.98	0.99993646	.015	4,235	9,073
.10	0.98	0.9998871	.016	7,587	5,721
.15	0.98	0.9998149	.020	9,360	3,948

Acronyms: W-F1=Weighted-F1.

In the next stage of the chained model, outlined in Section IV-B, the *confident* subset that filtered with an allowable error rate of 0.15 will be used as input. It is carefully chosen to ensure it is both reliable and large enough to effectively train and test the next classification model in the sequence. By using only high-confidence predictions, we aim to reduce errors and improve the performance of the next classifier. Allowing for a small margin of error in the *confident* subset could help the model generalize well on unseen data.

B. Fine-grained cyberbullying detection Model with and without the Proposed Filter Layer in the First Classification Phase

The results are shown in Table V, corresponding to weighted metrics and the CI metric. The comparison shows that filter

layer positively contributes to the classification score of the chain model, as the F1 score with the filter layer reaches 0.91, compared to 0.80 without it. This is because the *Confident subset* is used as input attributes during the training and prediction, whereas in the experiment without the filter layer, a subset of predicted labels ignoring the confidence estimation is used. These results demonstrate that, using the *Confident subset* to train the upper classifier in the chain offers a distinct advantage over the standard chain model.

TABLE V
RESULTS ON CYBY24 DATASET: THE PERFORMANCE OF THE CHAINED FINE-GRAINED CYBERBULLYING DETECTION FOR THE TEST SET, WITH AND WITHOUT THE PROPOSED FILTER LAYER.

Approach	**W-Pc±CI**	**W-Rc±CI**	**W-F1±CI**	**Test loss**	**Val. loss**
Proposed work using filter layer	**0.91±0.040**	**0.92±0.040**	**0.91±0.035**	0.18	0.18
Without filter layer	0.77±0.030	0.84±0.026	0.80±0.029	0.46	0.40

Acronyms: W-F1=Weighted-F1, W-Pc=Weighted-Precision, W-Rc=Weighted-Recall, val.=validation.

V. CONCLUSION

The traditional CC method suffers from error propagation, where misclassifications in earlier stages affect subsequent predictions. This study aims to enhance the predictive performance of the CC method by addressing the issue of error propagation. To this end, a filter layer-based technique is proposed. The filter layer acts as a mechanism to generate a confident subset from the predictions made in the first classification phase. This subset is then used in the second phase, helping to reduce the impact of earlier errors. Experiments were conducted on the CYBY24 dataset, where the proposed method achieved a best overall F1-score of 0.91, compared to 0.80 in the standard chain model. Future work will test the approach on new multi-label datasets and refine the architecture and uncertainty estimation techniques to improve reliability.

DATA AVAILABILITY

The corpus is useful for the research on leveraging bystander for cyberbullying detection, so it is made publicly available at https://www.kaggle.com/datasets/sllresearchgroup/cyber-bystander-role-labelled-dataset-cyby24

ACKNOWLEDGMENT

We thank the Ministry of Higher Education Malaysia (MOHE) and Universiti Sains Malaysia. This work is supported by the Fundamental Research Grant - FRGS/1/2023/ICT02/USM/02/1.

REFERENCES

[1] H. S. Alfurayj, S. L. Lutfi, and R. Perumal, "A Chained Deep Learning Model for Fine-grained Cyberbullying Detection with Bystander Dynamics." *IEEE Access*, vol. 12, no. August, pp. 105 588–105 604, 2024.

[2] N. K. Mishra and P. K. Singh, "Linear Ordering Problem based Classifier Chain using Genetic Algorithm for multi-label classification," *Applied Soft Computing*, vol. 117, no. January, 2022.

[3] J. Read, B. Pfahringer, G. Holmes, E. Frank, D. Xin, S. Takamichi, H. Saruwatari, W. Liu, and I. W. Tsang, "Classifier chains for multi-label classification," *Machine Learning*, vol. 85, no. 3, pp. 333–359, 2011. [Online]. Available: http://arxiv.org/abs/2206.10695

[4] S. C. Hunter, J. M. Boyle, and D. Warden, "Perceptions and correlates of peer-victimization and bullying," *British Journal of Educational Psychology*, vol. 77, no. 4, pp. 797–810, 2007.

[5] R. Kowalski, S. Limber, S. Limber, and P. Agatston, *Cyberbullying: Bullying in the digital age*. John Wiley & Sons, Reading, MA, 2012.

[6] R. M. Kowalski, G. W. Giumetti, A. N. Schroeder, and M. R. Lattanner, "Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth," *Psychological Bulletin*, vol. 140, no. 4, pp. 1073–1137, 2014.

[7] J. W. Patchin and S. Hinduja, *An update and synthesis of the research.Cyberbullying Prevention and Response: Expert Perspectives*, 2012.

[8] H. Li, H. An, Y. Wang, J. Huang, and X. Gao, "Evolutionary features of academic articles co-keyword network and keywords co-occurrence network: Based on two-mode affiliation network," *Physica A: Statistical Mechanics and its Applications*, vol. 450, pp. 657–669, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.physa.2016.01.017

[9] D. Xin, S. Takamichi, and H. Saruwatari, "Exploring the Effectiveness of Self-supervised Learning and Classifier Chains in Emotion Recognition of Nonverbal Vocalizations," *In Proc. ICML ExVo Workshop.*, 2022. [Online]. Available: http://arxiv.org/abs/2206.10695

[10] S. Radovanović, A. Petrović, B. Delibašić, and M. Suknović, "A fair classifier chain for multi-label bank marketing strategy classification," *International Transactions in Operational Research*, vol. 30, no. 3, pp. 1320–1339, 2023.

[11] T. Komatsu, S. Watanabe, K. Miyazaki, and T. Hayashi, "Acoustic event detection with classifier chains," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 1, pp. 46–50, 2021.

[12] J. Wang, K. Fu, and C. T. Lu, "SOSNet: A Graph Convolutional Network Approach to Fine-Grained Cyberbullying Detection," *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020*, pp. 1699–1708, 2020.

[13] J. Li, X. Zhu, and J. Wang, "AdaBoost. C2: boosting classifiers chains for multi-label classification." *In Proceedings of the AAAI Conference on Artificial Intelligence.*, vol. 37, no. 7, pp. 8580–8587, 2023.

[14] W. Liu and I. W. Tsang, "On the optimality of classifier chain for multi-label classification," *Advances in Neural Information Processing Systems*, vol. 2015-Janua, pp. 712–720, 2015.

[15] D. Jiaman, Z. Shujie, L. Runxin, F. Xiaodong, and J. Lianyin, "Association Rules-Based Classifier Chains Method," *IEEE Access*, vol. 10, pp. 18 210–18 221, 2022.

[16] R. Wang, S. Ye, K. Li, and S. Kwong, "Bayesian network based label correlation analysis for multi-label classifier chain," *Information Sciences*, vol. 554, pp. 256–275, 2021.

[17] A. Abdullahi, N. A. Samsudin, S. K. A. Khalid, and Z. A. Othman, "An Improved Multi-label Classifier Chain Method for Automated Text Classification," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 3, pp. 442–449, 2021.

[18] S. A. Azzi and C. B. O. Zribi, "A new Classifier Chain method of BERT Models For Multi-label Classification of Arabic Abusive Language on Social Media," *Procedia Computer Science*, vol. 225, pp. 476–485, 2023. [Online]. Available: https://doi.org/10.1016/j.procs.2023.10.032

[19] C. Atik, R. A. Kut, R. Yilmaz, and D. Birant, "Support Vector Machine Chains with a Novel Tournament Voting," *Electronics (Switzerland)*, vol. 12, no. 11, pp. 1–16, 2023.

[20] B. B. Jia and M. L. Zhang, "Decomposition-Based Classifier Chains for Multi-Dimensional Classification," *IEEE Transactions on Artificial Intelligence*, vol. 3, no. 2, pp. 176–191, 2022.

[21] S. N. Yildiz, F. Y. Okay, A. Islamov, and S. Özdemir, "Improved Chain-based Multi-Output Classification for Packaging Planning," *Procedia Computer Science*, vol. 231, no. 2023, pp. 697–702, 2024.

[22] Y. C. C. Alarcón and S. Destercke, "Multi-label Chaining with Imprecise Probabilities," *In European Conference on Symbolic and Quantitative Approaches with Uncertainty .*, vol. 12897 LNAI, no. Cham: Springer International Publishing., pp. 413–426, 2021.