

Deep Learning Approach for Multi-Class Identification from Domestic Violence Online Posts

Abhay Pravin Yadav¹, Jigar Vishanji Saraiya², Chaitanya Vasant Garud³, Shivam Rajendra Pate⁴ and Payel Thakur⁵

¹⁻⁵Dept. of Computer Engineering, Pillai College of Engineering, Panvel, 410206, Maharashtra, India
Email: ayadav20comp@student.mes.ac.in, jsaraiya20comp@student.mes.ac.in, chaitanya20comp@student.mes.ac.in, pshivam20comp@student.mes.ac.in, payelthakur@mes.ac.in

Abstract— Deep Learning for Multi-Class Identification From Domestic Violence Online Posts involves training a neural network to recognize patterns and features in text that are indicative of domestic violence. The aim is to develop, train and evaluate a multiclass classification model on a dataset of online posts so that model can identify and classify domestic violence-related content from online posts. By conducting various analysis tests, models will predict accurately and distinguish between different types of domestic violence, in order to better understand the prevalence and specific forms of abuse occurring within online communities. The model will be trained on a dataset of online posts and will be evaluated on its ability to accurately classify new, unseen data. The audio, selected text, image and will be classified to display the results. The outcome of this project will be a deep learning model that can effectively identify and classify different forms of domestic violence in online posts.

Index Terms— browser extensions, text, image, audio

I. INTRODUCTION

Fundamentals

1. Data collection and preprocessing:

- **Data Collection:** This phase involves gathering a substantial amount of online posts related to domestic violence. You can use web scraping tools to collect data from various sources such as social media platforms, forums, blogs, and news articles. It's essential to ensure that your dataset is diverse and representative of different types of content and platforms.

- **Data Preprocessing:** In this step, you clean and preprocess the collected data. This includes tasks such as text cleaning, removing HTML tags, handling special characters, and lowercasing the text.

2. Data splitting: Split the labeled dataset into training and sets.

3. Deep learning model selection: Select an appropriate deep learning model.

4. Model training: Select an appropriate deep learning architecture for your text, image and audio classification task. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), or more advanced models, depending on the complexity of the task and the size of your dataset. The choice of model may also depend on the specific characteristics of your text data.

5. *Model testing*: Finally, test the trained and validated deep learning model on the testing dataset to determine its performance on new, unseen data. Analyze the model's performance and fine-tune it if necessary. This step may involve adjusting the model architecture or retraining it with additional data if available.

II. MOTIVATION

The project on "Deep Learning for Multi-Class Identification From Domestic Violence Online Posts" aims to leverage neural network technology to discern patterns indicative of domestic violence within online content. By training a multiclass classification model on a dataset of online posts, the project seeks to accurately identify and classify various forms of domestic violence. Through rigorous analysis tests, the model will distinguish between different types of abuse, shedding light on prevalence and specific manifestations within online communities. The trained model will undergo evaluation on its capability to accurately classify new, unseen data encompassing audio, selected text, and images. Ultimately, the project endeavors to yield a powerful deep learning model capable of effectively identifying and classifying diverse forms of domestic violence present in online posts.

III. PROBLEM STATEMENT

Deep Learning for Multi-Class Identification From Domestic Violence Online Posts involves training a neural network to recognize patterns and features in text that are indicative of domestic violence. The aim is to develop, train and evaluate a multiclass classification model on dataset of online posts so that model can identify and classify domestic violence-related content from online posts. By conducting various analysis tests, models will predict accurately and distinguish between different types of domestic violence, in order to understand the prevalence and specific forms of abuse occurring within online communities. The model will be trained on a dataset of online posts and will be evaluated on its ability to accurately classify new, unseen data. The trained model will be integrated with developed browser extensions. The outcome will be a deep learning model that can effectively identify and classify different forms of domestic violence in online posts.

IV. OBJECTIVE

The objective of this work is as follows:

1. To develop a deep learning model that can identify and classify domestic violence-related content from online posts.
2. To train and evaluate the model on a dataset of online posts and conduct various analysis tests to ensure accurate predictions.
3. To distinguish between different types of domestic violence, in order to better understand their prevalence and specific forms of abuse.
4. To evaluate the model's ability to accurately classify new, unseen data using selected text and image classification techniques.
5. To produce a deep learning model that can effectively identify and classify different forms of domestic violence in online posts.

V. LITERATURE SURVEY

Iterative Semi-Open Coding, Ncapture:

In the research paper "Characterization of Domestic Violence through Self-disclosure in Social Media: A Case Study of the Time of COVID-19", the authors utilized two primary techniques to gain insights into how domestic violence was being experienced and self-disclosed on social media during the COVID-19 pandemic. The first technique was Iterative Semi-Open Coding, a qualitative research method that involves repeated cycles of coding and analysis. By iteratively coding and analyzing the self-disclosed experiences of domestic violence on social media, the researchers were able to identify the different types of domestic violence experiences that were being disclosed and the various coping mechanisms that were being employed by survivors. This approach allowed for a more flexible and nuanced analysis of the data, as the coding scheme could be adjusted to capture the emerging themes and patterns in the data.

The second technique utilized by the authors was Ncapture, a web-based data collection tool designed to capture social media data. By combining Ncapture with Iterative Semi-Open Coding, the authors were able to analyze the

data collected from social media in a more systematic and rigorous manner, providing a better understanding of how domestic violence was being experienced and self-disclosed on social media during the pandemic [3].

Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRUs):

The research paper titled "Deep Learning for Multi-Class Antisocial Behavior Identification From Twitter" utilizes a range of deep learning models, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRUs) to analyze different types of data and detect gender-based violence. CNNs are used to analyze image data, identifying visual patterns and features that may indicate violence. RNNs, LSTM networks, and GRUs are used to analyze sequential data, such as audio and text data, to identify patterns and indicators of violence [4].

Extract, Transform, and Load (ETL) tool, Inferential analysis technique, Descriptive analysis technique:

The research paper titled "A Multi-Modal Approach for Gender-Based Violence Detection" is a comprehensive study that employs a diverse set of analytical tools and techniques to address the problem of gender-based violence detection. The study utilizes deep learning algorithms to analyze a variety of data sources.

In addition to deep learning, the study also employs an Extract, Transform, and Load (ETL) tool to prepare and preprocess the data for analysis. The ETL tool is used to extract data from various sources, transform it into a structured format, and load it into a database for analysis. The research paper also utilizes inferential and descriptive analysis techniques to extract insights and draw conclusions from the data. Inferential analysis is used to make predictions and draw conclusions about the population based on a sample of data, while descriptive analysis is used to summarize and describe the characteristics of the data [5].

CNNs, RNNs (LSTMs and GRUs), BLSTM network:

In the research paper "Deep Learning for Multi-Class Identification From Domestic Violence Online Posts" the authors aimed to develop a multi-class classification model for identifying various types of domestic violence from online posts. To achieve this, they used a combination of deep learning techniques, including CNNs, RNNs (LSTMs and GRUs), and BLSTM networks. The dataset used in the study was sourced from various online platforms such as social media and online forums.

The authors utilized CNNs to extract essential features from the text data. Additionally, RNNs were used in the model to capture the temporal dependencies present in the text. LSTMs and 6 GRUs, popular RNN variants, were used to mitigate challenges related to vanishing and exploding gradients that can arise when processing lengthy sequences of text. The authors further utilized BLSTM networks to capture past and future inputs and provide additional context, thereby improving the model's performance [6].

CNN (ResNet - 152), LSTM:

In the research paper "Application of deep learning techniques: Domestic Violence Detection" [7], the authors aimed to develop a domestic violence detection model using deep learning techniques. To achieve this, they utilized a combination of CNNs and LSTM networks as the primary deep learning algorithms.

CNNs are primarily used in computer vision tasks, but they can also be applied to natural language processing (NLP) tasks such as text classification. In the authors' model, CNNs were utilized to extract features from the textual data. LSTM networks can effectively process and remember long sequences of text data, making them suitable for NLP tasks. By using LSTM networks, the authors were able to capture the sequential dependencies present in the text data, improving the model's accuracy in detecting domestic violence [7].

SN	Techniques	Author & Year of Publication	Advantages and Disadvantages
1.	Iterative Semi-Open Coding, Ncapture	Abdulrahman Aldkheel Lina Zhou Kanlun Wang 2021	Advantages: The project respects the privacy of victims, as they self-disclose their experience of domestic violence voluntarily and the project can analyze a large sample of social media posts, allowing for a comprehensive understanding of the scope and nature of domestic violence during COVID-19. Disadvantages: In this paper the sample size which was taken for the content analysis is very small. Also this project's analysis was about the self-disclosed DV experience but

			with the support of the comments of the social media posts can also help to find the overall community responses.
2.	Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRUs)	Ravinder Singh Sudha Subramani Jiahua Du Yanchun Zhang Hua Wang Khandakar Ahmed Zhenxiang Chen 2020	Advantages: The project provides evidence-based data on the prevalence, types, and dynamics of antisocial behavior on Twitter, which can inform policy and practice. Disadvantages: The study does not discuss the potential biases and fairness issues associated with the development and deployment of the model.
4.	CNNs, RNNs (LSTMs and GRUs), BLSTM network	Sudha Subramani Sandra Michalska Hua Wang Jiahua Du 2019	Advantages: Deep Learning models with GloVe embedding achieved improved performance over the traditional Machine Learning classifiers except for RNNs. Disadvantages: Lack of research on the application of deep learning models for multi-class identification of different types of abuse (e.g., physical, emotional, sexual) in online posts related to domestic violence.
5.	XGBoost	Md. Rokonzaman Reza Fabiha Mukarrama Binte Mannan Dhrubo Barua	Advantages: Centralized Data Repository: The proposed system provides a centralized platform to store data on violence against women and children. This centralization can enhance data
		Shafayetul Islam 2021	accessibility and analysis for various stakeholders. Machine Learning for Prediction: By utilizing machine learning techniques, such as XG Boost, the system can estimate future violence trends. This predictive capability could aid in early intervention and prevention efforts. Disadvantages: Data Privacy Concerns: Storing sensitive information about victims in a database raises significant privacy and security concerns. Ensuring data protection and compliance with privacy regulations is crucial. Bias in Machine Learning Models: Machine learning models can inherit biases from the data they are trained on. It's important to ensure that the models do not perpetuate existing biases in violence data.
6.	Robotics Machine learning	Nina Sunde Inger Marie Sunde 2021	Advantages: Clear Objective: The summary clearly outlines the research paper's overarching objective, which is to develop and deploy an AI-based police robot for preventing online child sexual exploitation and abuse (CSEA). Multidisciplinary Approach: The summary highlights that the research takes a multifaceted approach by considering aspects of technology, criminology, and legal analysis. This multidisciplinary approach can provide a comprehensive view of the issue. Disadvantages: Lack of Specifics: The summary provides a general overview of the research's objectives and approach but lacks specific details regarding the methodologies, technologies, or legal considerations involved in developing and deploying PrevBOT.

			Absence of Key Findings: The summary does not mention any specific findings or results from the research, making it challenging to assess the practical implications or success of the PrevBOT concept.
--	--	--	---

VI. PROPOSED SYSTEM

A. Overview

1) Data Preprocessing:

In this step, the initial dataset is subjected to feature engineering and feature selection. This involves cleaning, transforming, and refining the dataset to extract meaningful features that will be used for model training. Feature engineering may include text preprocessing, feature extraction.

2) Data Splitting

After the dataset is prepared, it is split into three subsets: training, testing, and validation. The training dataset is used to teach the predictive model, the testing dataset is used to assess its performance, and the validation dataset is used to fine-tune the model and prevent overfitting.

3) Model Creation

In this stage, a predictive model is developed. This model is designed to predict instances of domestic violence in text, image, and audio data. Depending on the specific types of data and the complexity of the problem, various machine learning or deep learning algorithms are employed.

4) Model Validation

The trained model is evaluated using the validation dataset. This step ensures that the model's performance is acceptable and helps in making adjustments if necessary. It is crucial to monitor the model's behavior on a separate dataset to prevent overfitting.

5) Model Testing

Once the model has successfully performed on the validation dataset, it is then tested on the independent testing dataset. This testing step verifies that the model generalizes well to unseen data and that its predictive capabilities meet the project requirements.

6) Deployment as a Chrome Extension

If the model passes the testing phase, it can be deployed as a Chrome extension. This extension can be integrated into a web browser and used to predict domestic violence in real-time for text, image, and audio data.

7) Model Extension Evaluation

The Chrome extension is used to evaluate real-world data, which can be in the form of text, images, or audio. This step assesses how well the model extension performs when applied to actual data from users' web browsing activities.

8) Displaying Results

The results of the domestic violence predictions made by the Chrome extension are displayed as pop-ups within the user's browser. This allows users to be informed about potential instances of domestic violence as they browse the web.

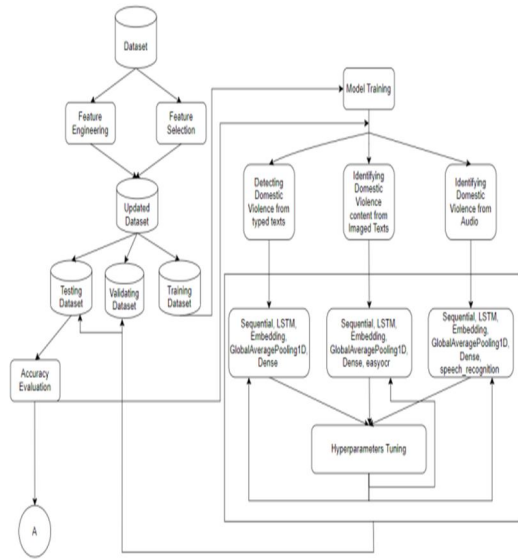
• Detecting Domestic Violence from Comments

1) Sequential Model

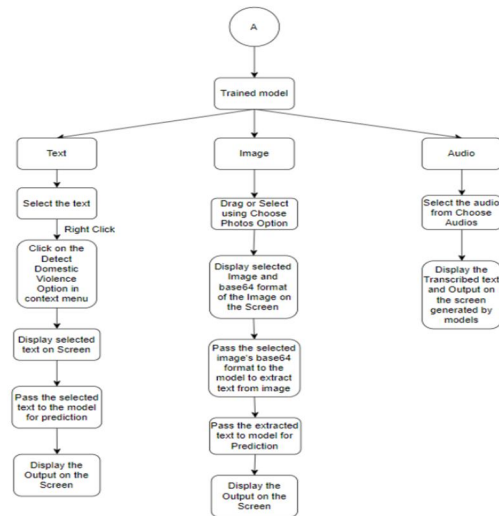
This is the foundational structure for creating a neural network in Keras. It allows you to build a model layer by layer.

2) LSTM

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) that can be particularly useful in identifying text. LSTMs are designed to overcome the vanishing gradient problem that can occur in traditional RNNs, allowing them to capture long-term dependencies in sequential data, such as text.



Proposed system architecture (Part - 1)



Proposed system architecture (Part - 2)

3) Embedding Layer

This layer is used for word embedding in Natural Language Processing (NLP) tasks.

4) GlobalAveragePooling1D Layer

This layer computes the average of the output sequence from the LSTM layer for each input sequence. It reduces the sequence data to a fixed-length vector, which can be useful for classification tasks.

• Detecting Domestic Violence from Images

1) Sequential Model

This is the foundational structure for creating a neural network in Keras. It allows you to build a model layer by layer.

2) LSTM

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) that can be particularly useful in identifying text. LSTMs are designed to overcome the vanishing gradient problem that can occur in traditional RNNs, allowing them to capture long-term dependencies in sequential data, such as text.

3) Embedding Layer

This layer is used for word embedding in Natural Language Processing (NLP) tasks.

4) GlobalAveragePooling1D Layer

This layer computes the average of the output sequence from the LSTM layer for each input sequence. It reduces the sequence data to a fixed-length vector, which can be useful for classification tasks.

5) easyocr

easyocr is a Python library for Optical Character Recognition (OCR), which is a technology used to extract text and characters from images or scanned documents.

• Detecting Domestic Violence from Audio

1) Sequential Model

This is the foundational structure for creating a neural network in Keras. It allows you to build a model layer by layer.

2) LSTM

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) that can be particularly useful in identifying text. LSTMs are designed to overcome the vanishing gradient problem that can occur in traditional RNNs, allowing them to capture long-term dependencies in sequential data, such as text.

3) Embedding Layer

This layer is used for word embedding in Natural Language Processing (NLP) tasks.

4) GlobalAveragePooling1D Layer

This layer computes the average of the output sequence from the LSTM layer for each input sequence. It reduces the sequence data to a fixed-length vector, which can be useful for classification tasks.

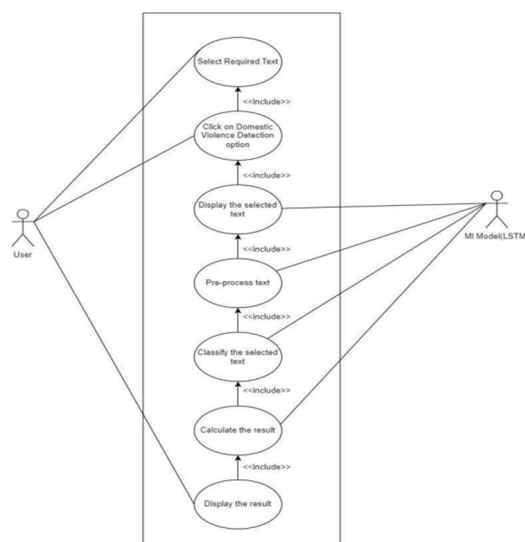
5) speech_recognition

speech_recognition model is used to convert spoken language into text.

Dataset:

Dataset	Size
jigsaw-unintended-bias-in-toxicity-classification	1971916

B. Use Case Diagram



Use Case of Domestic Violence Detection

- **Text Selection:** The user selects the text they want to analyze. This could be any form of text that the user suspects contains instances of domestic violence.
- **Option Selection:** The user clicks on the domestic violence detection option. This action triggers the machine learning model to start analyzing the selected text.
- **Display Selected Text:** The model displays the selected text. This step is crucial for the user to verify that the correct text has been selected for analysis.
- **Text Pre-Processing:** The model pre-processes the text. This step involves cleaning and formatting the text so that it can be effectively analyzed by the machine learning model.
- **Text Classification:** The model classifies the text. Using trained algorithms, the model identifies patterns and features in the text that are indicative of domestic violence.
- **Result Calculation:** The model calculates the result. Based on the classification, the model determines whether or not the text contains instances of domestic violence.
- **Display Result:** The model displays the result to the user. The result could be a simple binary output (domestic violence present or not), or a more detailed report highlighting specific instances of domestic violence in the text.

C. Hardware and Software Specifications

The experiment setup is carried out on a computer system which has the different hardware and software specifications as given in Table (a) and Table (b) respectively.

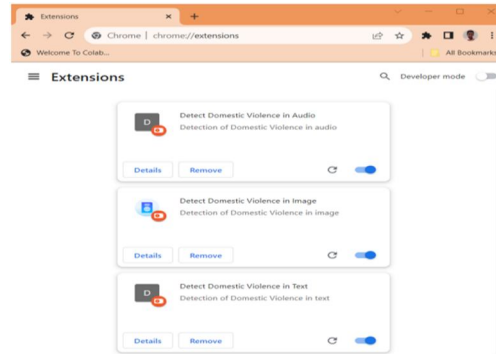
Processor	2 GHz Intel
HDD	180 GB
RAM	2 GB

(a)

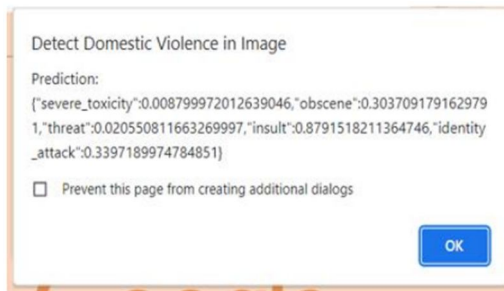
Operating System	Windows XP Professional With Service pack 2
Programming Language	HTML, CSS, JavaScript, Python

(b)

V. RESULT



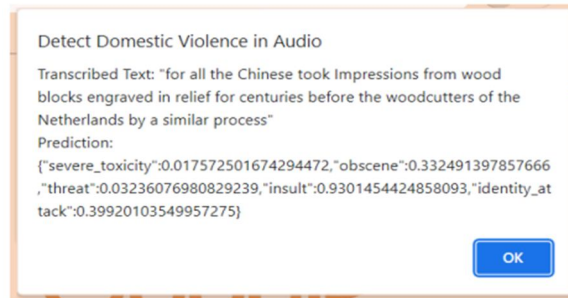
Inputs



Output (Prediction for text)



Output (Prediction for image)



Output (Prediction for audio)

VI. CONCLUSION

In this report, the study of different domain techniques is presented. It delves into the intricacies of distinct algorithms and methods, including deep learning models such as Sequential, LSTM, Embedding, GlobalAveragePooling1D, and Dense, alongside natural language processing tools like easyocr and speech_recognition. The dataset identified for experiments is jigsaw-unintended-bias-in-toxicity-classification. The report delves into the practical application of these algorithms and their effectiveness in addressing real-world challenges, particularly in the context of understanding and predicting toxicity in provided content.

ACKNOWLEDGMENT

We express our sincere gratitude to Mrs. Payel Thakur, our project guide, and the Head of the Department for their priceless support and guidance during the successful execution of our major project in AI and ML. Their significant contributions have played a pivotal role in our academic achievements, and we owe them a debt of gratitude for their mentorship and unwavering dedication.

REFERENCES

- [1] Reza, Md. Rokonzaman & Mannan, Fahiha & Barua, Dhruvo & Islam, Shafayetul & Imtiaz Khan, Nafiz & Mahmud, Sharifa. (2021). Developing a Machine Learning Based Support System for Mitigating the Suppression Against Women and Children. 1-6.10.1109/ICEEICT 53905.2021.9667924.
- [2] Sunde, Nina & Sunde, Inger Marie. (2021). Conceptualizing an AI-based Police Robot for Preventing Online Child Sexual Exploitation and Abuse: Part I – The Theoretical and Technical Foundations for PrevBOT. Nordisk Politiforskning.2. 1-21. 10.18261/issn.2703-7045-2021-02-01.
- [3] Aldkheel, Abdulrahman & Zhou, Lina & Wang, Kanlun. (2021). Characterization of Domestic Violence through Self-disclosure in Social Media: A Case Study of the Time of COVID-19.10.1109/ISI53945.2021.9624676.
- [4] Singh, Ravinder & Subramani, Sudha & Du, Jiahua & Zhang, Yanchun & Wang, Hua & Ahmed, Khandakar & Chen, Zhenxiang. (2020). Deep Learning for Multi-Class Antisocial Behavior Identification From Twitter. IEEE Access. 8. 194027-194044. 10.1109/ACCESS.2020.3030621.
- [5] Khatri, Hamida & Abdellatif, Iheb. (2020). A Multi-Modal Approach for Gender-Based Violence Detection. 144-149. 10.1109/IEEECloudSummit48914.2020.00028.
- [6] S. Subramani, S. Michalska, H. Wang, J. Du, Y. Zhang and H. Shakeel, "Deep Learning for Multi-Class Identification From Domestic Violence Online Posts," in IEEE Access, vol. 7, pp. 46210-46224, 2019, doi: 10.1109/ACCESS.2019.2908827.
- [7] Haque, A K M Bahalul. (2019). Application of deep learning techniques: Domestic Violence Detection.
- [8] UN Global Dataset on Violence Against Women. (2020). Retrieved on August 10, 2020 from <http://evaw-global-database.unwomen.org/en>
- [9] Grogger, J., Ivandic, R., & Kirchmaier, T. (2020). Comparing Convolutional and Machine Learning Approaches to Risk Assessment in Domestic Abuse Cases. Center for Economics Performance.
- [10] Aakervik, A-L. (2020, 2 November). AIBA– Avslørercybergrooming.<https://ntnudiscovery.no/aibaavslorer-cybergrooming/>.
- [11] Colorado State University. (2019). Exposure to Air Pollution Increase Violent Crime Rates.ScienceDaily.
- [12] World Health Organization. (2020). Violence by Intimate Partners.