JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

GENDER-BASED VIOLENCE DETECTION USING MACHINE LEARNING

Saamya Gupta, Rutuparn Kakade, Sahil Akalwadi, Soham Pawarm, Anurag Tekam
Student
Vishwakarma Institute of Technology, Pune

***Abstract* — Gender Based Violence is deep rooted social issue in today's world. Even in todays advanced world, Gender based violence continues to be major issue which affects survivors as well families leading to social disputes and interrupts social harmony. There are various types of gender based violence. We are addressing physical gender based violence and gender based violence through text. Our proposed system basically prevents hate speech in online social media platform thus moderating social media contents. Second part of our proposed system is based on ML. We have designed gender based violence detection model which is based on pre trained I3D action recognition model, it basically identifies frame by frame gender first and proceeds to process the frame, if violence is detected it interprets it as gender based violence and further sends an alert text message.**

Keywords* — *Gender based violence, GBV model, hate speech prevention, Machine Learning

I. INTRODUCTION

The United Nations' definition of GBV is, "any act of gender-based violence that results in, or is likely to result in, physical, sexual or psychological harm or suffering to women…whether occurring in public or private life." A broad definition of gender-based violence (GBV) is any violence that has its origins in social roles and unequal power structures. Being overrepresented among individuals with less power increases the likelihood that women and children may encounter GBV. However, study among male abusers found GBV is severely underrepresented, particularly in populations affected by conflicts. For women, the definition of GBV is typically expressed in terms of violence against women, where it is more precisely defined.

The effects of GBV are widespread, having an impact on the survivors' health and well-being as well as that of their families and societies. Research shows following results in terms of gbv. A little more than 1 in 3 women had been the victim of violence from a spouse or family member. Nearly 40% of murdered women are killed by their partners. Women make up 90% of rape victims.

Over 90% of young victim victims and 80% of adult victims both know their rapists. Many women believe that sometimes violence towards wives is acceptable. In other nations, such as Ethiopia, India, Bhutan, Samoa, and Laos, more than 50% of women share this opinion. More than 90% of criminals never receive justice. Gender based violence is of different type depending on the situation. In our project we have addressed the two types that are physical violence and text speech abuse. For many women, hearing sexist hate speech is a regular occurrence, and it is only now that it is being addressed. Fighting gender stereotypes needs both action and legal remedies.

Expressions that transmit, provoke, advocate, or otherwise justify sex-based hatred are referred to as sexist hate speech. Young women, women in the media, and women in politics are some of the groups that are specifically targeted by sexist hate speech, but every woman and girl is a potential target for both online and offline sexist hate speech. The prevalence of sexist hate speech has increased as a result of the Internet's expanding accessibility and popularity.

II. PROPOSED SYSTEM

In the figure 1, we have illustrated our system architecture for proposed system. There are three main functionalities of our project. First we have get helpline. We simply enter the name of our state. And it will redirect to web url of corresponding to that state's emergency contact. Second is content moderation in social media which is used to prevent hate speech in text. Third and main

functionality of our project is gender based violence detection, basically we have used pretrained model I3D which is action detection model. The model is trained on kinematics data set. We have modified and used transfer learning approach by adding violence and non violence data set to the model, so that the model can detect violence. Further we have implanted gender recognition algorithms through opencv/openface so that it recognize and separated gender based violence detection.

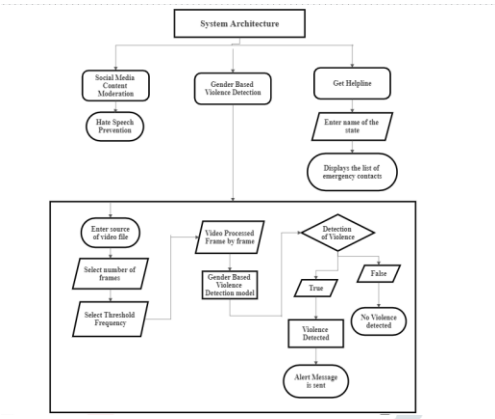

Fig1. Proposed System

III. LITERATURE REVIEW

Physical: It causes illnesses, distress, and injury. Physical abuse commonly takes the form of beating, strangulation, pushing, and the use of weapons. Since the age of 15, 31% of women in the EU had been the victim of one or more physical assaults.

Sexual: encompasses activities that are directed against someone's sexuality without that person's consent. It also includes attempts to elicit sexual favours. According to estimates, one in 20 women (5%) in EU nations had experienced a sexual assault since turning 15.

Blackmail, control, coercion, and other psychologically abusive behaviors are included under the heading "psychological." In the 28 EU member states, 43% of women have been the victim of psychological abuse by a romantic relationship. A variety of methods have been explored for the hate speech detection task, including traditional classifiers, deep learning-based classifiers , or the combination of both approaches . Classifiers like support vector machines (SVM), extreme gradient boosting (XGB), and multi-layer perceptron (MLP) are commonly used in this task, which typically require vector representations of the text data. the aim of paper is to provide a thorough empirical evaluation and comparison of different types of hate speech detection methods on these datasets.

With the help of existing criminological theories and literature this paper identifies social and situational policy frameworks that could be adopted in preventing GBV in India. The case study approach used in this paper thus provides a comprehensive understanding of gender-based violence in India. Given the difficulty in applying optical flow-based human action recognition due to the significant amount of calculation required, an I3D-shufflenet model for human action recognition is proposed, combining the benefits of I3D neural networks and lightweight shufflenet models. based on a trained I3D-shufflenet model, human action is recognized and categorized. According to the experimental findings, the shuffle layer enhances the feature composition in each channel, which can encourage the usage of useful information.

The spatial-temporal properties of the object are retrieved using the Histogram of Oriented Gradients (HOG) for training, which can considerably increase the ability of human action expression and decrease the calculation required for feature extraction. On the UCF101 dataset, the I3D-shufflenet is put to the test and contrasted with other models. The ultimate outcome demonstrates that the I3D-shufflenet, with an accuracy of 96.4%, is more accurate than the original I3D.

IV. METHODOLOGY

A. Research

We researched and reviewed multiple research papers, blogs , articles on the topic gender based violence to get deep understanding of our problem statement. It helped us to understand and address which problem we can work on. After in depth analyzing gender based violence and types of GBV, we decided to work on physical gender based violence detection through machine learning approach and hate speech prevention through social media content moderation.

B. Tools & Technologies

1. Flutter

Flutter is an open-source UI software development kit created by Google. We have used flutter to develop frame integral framework of our project.

2. Dart Programming Language

Dart is a programming language designed for client development, such as for the web and mobile apps. It is developed by Google and can also be used to build server and desktop applications. It is an object-oriented, class-based, garbage-collected language with C-style syntax. We have used dart language for backend of our project

3. Machine Learning

Machine learning is a subfield of artificial intelligence, which is broadly defined as the capability of a machine to imitate intelligent human behavior. We have used machine learning approach to detect physical gender based violence.

4. Tensorflow

TensorFlow is a free and open-source software library for machine learning and artificial intelligence. TensorFlow is an end-to-end open source platform for machine learning.

5. Python Programming Language

Python is a vast programming language, used for various different purposes. In our project we used Python to implement Machine Learning.

6. Fast2sms

SMS Gateway API. Fast2sms SMS API work with PHP, JAVA, C#, C, Python, Ruby, JavaScript, NodeJS, etc. It is Secure, robust and easy to integrate APIs to send DLT Approved SMS.

We have used fast2sms API to get alert messages on detecting gender based violence.

7. Jinja

Jinja is a web template engine for the Python programming language. With Jinja, we have built rich templates that power the front end of our Python web application.

C. Machine learning model: For our project, we researched and analyzed different machine learning models and data sets. We compared their accuracy in terms of real time application. After viewing results, we came to conclude there was no trained gender based violence model which can detect gender based violence in the video. So we decided to design GBV model. First step was to select an action recognition model, for this we choose I3D action recognition model as it was more accurate on result test data sets UCF-101 and HMDB-101 as you can see in the figure comparing different action recognition models.The I3D model was presented by researchers from DeepMind and the University of Oxford in a paper called "Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset".

Model	UCF-101	HMDB-51
Two-Stream [25]	88.0	59.4
IDT [30]	86.4	61.7
Dynamic Image Networks + IDT [2]	89.1	65.2
TDD + IDT [31]	91.5	65.9
Two-Stream Fusion + IDT [8]	93.5	69.2
Temporal Segment Networks [32]	94.2	69.4
ST-ResNet + IDT [7]	94.6	70.3
Deep Networks [15], Sports 1M pre-training	65.2	-
C3D one network [29], Sports 1M pre-training	82.3	-
C3D ensemble [29], Sports 1M pre-training	85.2	-
C3D ensemble + IDT [29], Sports 1M pre-training	90.1	-
RGB-I3D, miniKinetics pre-training	91.8	66.4
RGB-I3D, Kinetics pre-training	95.4	74.5
Flow-I3D, miniKinetics pre-training	94.7	72.4
Flow-I3D, Kinetics pre-training	95.4	74.6
Two-Stream I3D, miniKinetics pre-training	96.9	76.3
Two-Stream I3D, Kinetics pre-training	**97.9**	**80.2**

fig 2. Comparison of different action recognition models. [17]

D. I3D model:- Two-Stream Inflated 3D ConvNet (I3D) that is based on 2D ConvNet inflation: filters and pooling kernels of very deep image classification ConvNets are expanded into 3D, making it possible
to learn seamless spatio-temporal feature extractors from video while leveraging successful ImageNet architecture designs and even their parameters. pre-training on Kinetics, I3D models considerably improve
upon the state-of-the-art in action classification, reaching 80.9% on HMDB-51 and 98.0% on UCF-101In simple terms, the architecture of inflated 3D CNN model goes something like this (fig3) – input is a video, 3D input as in 2-dimensional frame with time as the third dimension. It contains Convolutional(CNN) layers with stride 2, after which there is a max-pooling layer and multiple Inception modules (conv. Layers with one max pooling layer, concatenation is the main task). Inflated because of the reason that we are having these modules (described in the paper) dilated into the middle of the model. These modules can have different mini architectures in them like LSTM, Two streams and so on are mentioned in the paper. In the end, we have an average pooling layer with a 1x1x1 Conv layer for prediction.

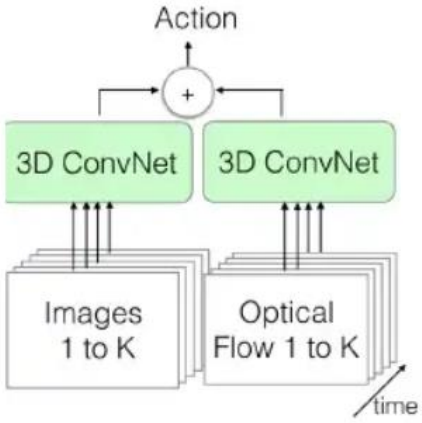

fig 3: Proposed Inflated 3D ConvNet (I3D)

E. The Kinetics Human Action Video Dataset:- In this dataset, the list of action classes covers: Person Actions (singular), e.g. drawing, drinking, laughing, punching; Person-Person Actions, e.g. hugging, kissing, shaking hands; and, Person-Object Actions, e.g. opening presents, mowing lawn, washing dishes.
The dataset has 400 human action classes, with 400 or more clips for each class, each from a unique video. The clips last around 10s, and there are no untrimmed videos.
The test set consists of 100 clips for each class. Therefore, making this data set vast and precise for training model.

V. Future Scope

In Our proposed system can be improved and can be made for reliable and practical to use by adding various features. In helpline contact, we can get access to users location and according to location, we can display the emergency contacts of that location by verifying where user is at. This will save time as user will have to just click on get helpline button and he/she will get access to all nearby emergency contacts. In our proposed system, we have added an simple social media content moderation algorithm which can prevent hate speech in social media. This can be further improved by reporting and blacklisting the user, if he/she continues to hate speech on regular basis. The details and text abuse evidences of that user thus can be further sent to authorities for further action.

Our gender based violence detection works on user input files, We can make more practical and real time system by adding live detection algorithm which can be used in cctvs to detect gender based violence. We can also improve the system by real time capturing of particular detection and forwarding that video clip with its information such as date, time, location to local authorities so that authorities can take quick action.

VI. Conclusion

Gender based violence continues to be a major issue throughout the world. Many of gender-based violence cases get neglected and unreported due to lack of confidence and resources. We have addressed this problem in our project and proposed a

gender-based violence detection system. Our System has two main features: Social media Content moderation that is to prevent hate speech in social media platforms. It prevents hate speech and curse words to be used in social media.

Second is gender-based violence detection model which detects types of physical violence in video frame by frame. We have used flutter framework to build our project and dart language as backend programming language. We have also used python and tensor flow for machine learning model. We have implemented I3D kinematics which is action detection pre trained model, and further improved this model by providing violence detection data sets and gender recognition algorithm to design our gender-based violence detection model. Our model basically can recognize violence based on gender. We have also used to fast2sms API to alert when gender-based violence is detected.

VII. ACKNOWLEDGMENT

We thank Professor Ranjana Jadhav for their guidance.

VIII. REFERENCES

[1] Malik, Jitendra Singh, Guansong Pang, and Anton van den Hengel. "Deep Learning for Hate Speech Detection: A Comparative Study." arXiv preprint arXiv:2202.09517 (2022).

[2] Saeed, Ari M., et al. "Hate Speech Detection in Social Media for the Kurdish Language." The International Conference on Innovations in Computing Research. Springer, Cham, 2022.

[3] Honarjoo, Narges, Ali Abdari, and Azadeh Mansouri. "Violence detection using pre-trained models." 2021 5th International Conference on Pattern Recognition and Image Analysis (IPRIA). IEEE, 2021.

[4] Rodríguez-Rodríguez, Ignacio, et al. "Modeling and forecasting gender-based violence through machine learning techniques." Applied Sciences 10.22 (2020): 8244.

[5] Babu, Dhanya. Gender Based Violence in India: An Analysis of National Level Data for Theory, Research and Prevention. City University of New York John Jay College of Criminal Justice, 2019.

[6] Simister, John Gordon. "Gender-based violence is a growing problem in India." Medical research archives 6.1 (2018).

[7] Haizhong, Qian. "I3D: An Improved Three-Dimensional CNN Model on Hyperspectral Remote Sensing Image Classification." Security and Communication Networks 2021 (2021).

[8] Carreira, Joao, and Andrew Zisserman. "Quo vadis, action recognition? a new model and the kinetics dataset." proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[9] Liu, Guocheng, et al. "I3d-shufflenet based human action Recognition." Algorithms 13.11 (2020): 301.

[10] Elias, Stanley, and Nubar Gurbanova. "Relocating gender stereotypes online: Critical analysis of sexist hate speech in selected social media." International Conference on Language Phenomena in Multimodal Communication (KLUA 2018). Atlantis Press, 2018.

[11] Russo, Nancy Felipe, and Angela Pirlott. "Gender-based violence: concepts, methods, and findings." Annals of the new york academy of sciences 1087.1 (2006): 178-205.

[12] Clark, Cari. "Gender-based violence research initiatives in refugee, internally displaced, and post-conflict settings: Lessons learned." (2003).

[13] Sumon, Shakil Ahmed, et al. "Violence detection by pretrained modules with different deep learning approaches." Vietnam Journal of Computer Science 7.01 (2020): 19-40.

[14] Mumtaz, Nadia, et al. "An overview of violence detection techniques: current challenges and future directions." Artificial intelligence review (2022): 1-26.

[15] John, Neetu, et al. "Lessons never learned: crisis and gender-based violence." Developing world bioethics 20.2 (2020): 65-68.

[16] https://github.com/deepmind/kinetics-i3d

[17] https://medium.com/nerd-for-tech/review-quo-vadis-action-recognition-a-new-model-and-the-kinetics-dataset-video-classification-a7535aa8bf48

[18] https://www.youtube.com/watch?v=Zchx0hPtLYE

[19] https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/gender-based-violence/what-gender-based-violence_en

[20] https://docs.fast2sms.com/#get-method6

[21] https://www.coe.int/en/web/gender-matters/types-of-gender-based-violence

[22] https://docs.flutter.dev/

[23] https://dart.dev/guides

[24] https://jinja.palletsprojects.com/en/3.1.x/