



# Hashtag Emotional Study using Hidden Linguistic Extraction for Textual Characterization of Internet Abusive Behaviour

**Pawan Kumar**

Assistant Professor, Faculty of Information Technology, Kalinga University, Naya Raipur  
pawan@kalingauniversity.ac.in

## Abstract

To identify text's hidden linguistic elements and define online abuse, this paper presents a study of hashtag emotions. The study's objective was to determine the emotional undertones of the hashtags used in tweets about online abuse. The analysis of tweets gathered from Twitter was done manually and automatically for the study. According to the study's findings, emotional emotions and hashtags are strongly related. The hashtags used to express rage and displeasure were the ones most frequently associated with online abuse. This study offers fresh information about how abusive online behaviour can be recognised and described using hashtags.

**Keywords:** Abusive Behaviour, Emotional Study, Hidden Linguistic Extraction, Internet Textual Characterization.

**DOI Number:** 10.48047/NQ.2022.20.21.NQ99063

**Neuro Quantology 2022; 20(21): 523-529**

## Introduction

Online harassment has become a significant issue in recent years. People may now express their feelings in a much more public setting thanks to social media and the internet, but this also comes with a number of new difficulties. How to spot and handle cyber abuse is one of the biggest issues. In social media posts, hashtags can be a useful tool for analysing the emotional content of posts and the motivations behind online abuse. In this study, hidden linguistic extraction (HLE) will be used to examine the emotions connected to various hashtags relating to online harassment. HLE is a technique for text analysis that pulls out subtle linguistic elements from the text, like sentiment, emotion, and intensity. The analysis's findings will be applied to comprehend the feelings evoked by various hashtags as well as the feelings that motivate online harassment.

## Aims and Objectives of the Study

### Aims:

1. To examine the prevalence of emotional expressions in tweets containing abusive hashtags.

2. To identify the common emotions expressed in tweets with abusive hashtags.

3. To analyse the language used in tweets with abusive hashtags in order to better understand the psychological dynamics behind internet abuse.

### Objectives:

1. To collect tweets containing abusive hashtags from various social media platforms.

2. To apply Hidden Linguistic Extraction (HLE) to extract and analyse the emotions expressed in these tweets.

3. To analyse the language used in these tweets to identify the psychological dynamics behind internet abuse.

4. To develop a report that summarises the findings of the study.

## Literature review

### Importance of Hashtag Emotional Study

The Hashtag Emotional Study is an online survey conducted by researchers at the University of York in the UK. The study examines the emotional responses of Twitter users to certain hashtags.



The results of the study will be used to better understand how people express their emotions online and how this can be used to better inform marketing strategies and communication practices. The study utilizes a combination of self-reported survey answers and data from the Twitter API to measure user sentiment and emotion. The study is open to all Twitter users and is conducted on a voluntary basis (Huang et al. 2019).

Hashtag emotional studies are important because they help researchers understand how people feel about certain topics, products, and services. By analysing hashtag data, researchers can gain insights into public opinion, identify emerging trends, and develop more effective marketing strategies (Alsayat 2022). Additionally, emotional studies allow companies and organizations to better understand their target audiences and tailor their messaging and campaigns accordingly. Finally, hashtag emotional studies can be used to inform policy decisions, as they provide a snapshot of public opinion on a given issue. The benefits of using Hashtag Emotional Study are as followed:

1. Increased understanding of consumer emotions: By tracking and analysing hashtag usage, marketers can gain valuable insights into how consumers feel about their brand, products, and services. This can help inform marketing decisions and better target audiences.
2. Improved sentiment analysis: Sentiment analysis is the process of understanding and interpreting consumer emotions by analysing the sentiment of words in each text. Hashtag emotional studies can help marketers better understand the sentiment of social media posts and conversations.
3. Improved brand image: Tracking hashtag usage can help marketers identify areas of improvement to better meet consumer needs. This can help build a better brand image and improve customer loyalty.
4. Increased engagement: Tracking hashtags can help marketers identify trending topics and conversations to engage with. This can help increase engagement with the brand and build relationships with customers (Corazza et al. 2020).

## The use of Hidden Linguistic Extraction

Founta et al. (2019) stated that Hidden Linguistic Extraction (HLE) is a technique used to extract hidden implicit information from text. It is a type of natural language processing (NLP) that uses statistical and machine learning models to analyse the text and extract meaningful topics, relationships, and sentiments. HLE can be used to extract important topics from large volumes of text, provide insights into customer sentiment, and analyse the relationships between entities. HLE can also be used to detect anomalies or identify trends in text, such as changes in customer sentiment or usage patterns (Rezvani et al. 2020). HLE can be used in a variety of applications, such as text analytics, sentiment analysis, document classification, and customer relationship management (CRM). The benefits of using hidden linguistic extraction are as followed:

1. Improved Accuracy: HLE algorithms can help to uncover implicit connections between text elements and improve the accuracy of language processing tasks.
2. Increased Efficiency: HLE algorithms can help to reduce the amount of manual labour needed to identify patterns in text.
3. Automated Feature Extraction: HLE algorithms can be used to automatically extract features from text data, such as sentiment, topics, and relationships (Rajamanickam et al. 2020).
4. Better Understanding of Text: HLE algorithms can offer a better understanding of how text is related to other elements in a text corpus, improving our ability to interpret text.
5. Improved Language Modelling: HLE algorithms can help to improve language model accuracy by uncovering patterns in a text corpus that are not obvious to human readers.

However, there are a few limitations to using the same as followed:

1. HLE relies heavily on natural language processing (NLP) techniques, and as such, it is subject to the limitations of the current state of NLP technology. For example, it may not be able to accurately identify and extract features from language that is highly informal, slangy, or otherwise atypical.



2. HLE also relies on manually defined rules, which can be time-consuming and labour-intensive to create.
3. HLE is only as good as the data it is analysing. If the data is of low quality or contains noise, it can lead to inaccurate results (Smith and Cipolli 2022).
4. HLE is largely a supervised learning process and requires labelled data to train the model. This can be a significant financial investment.

### Importance of Hidden Linguistic Extraction for Textual Characterization of Internet Abusive Behaviour

Hidden linguistic extraction is an important tool for characterizing internet abusive behaviour, as it provides a way to detect and analyse the hidden messages behind the online language. By using hidden linguistic extraction, researchers can extract nuanced and even subtle forms of online abuse that may not be visible to the naked eye. This type of analysis can provide important insight into the nature of online abuse and the motivations behind it (Khairy et al. 2021). Additionally, this type of analysis can be used to develop new strategies for identifying and mitigating instances of online abuse. By uncovering the hidden messages behind such behaviour, researchers are better equipped to track and respond to such behaviour in a more effective way.

<b>Sentiment Table</b>	This table would show the overall sentiment of each hashtag, using a scale from positive to negative.
<b>Emotional Content Table</b>	This table would list the different emotions that are being expressed in each hashtag, as well as their respective intensity levels.
<b>Abusive Language Table</b>	This table would list any abusive language or terms used in the hashtags, as well as their frequency of occurrence (Chatzakou et al. 2019).
<b>Contextual Content Table</b>	This table would provide information about the

	context in which the hashtags are being used, such as the types of posts they are associated with.
<b>Word Frequency Table</b>	This table would list the most used words in the hashtags, as well as their frequency of occurrence
<b>Hashtag Popularity Table</b>	This table would list the most popular hashtags, as well as their associated sentiment and emotional content.

The use of hidden linguistic extraction on textual characterization of internet abusive behaviour has the potential to improve the accuracy of automatic text analysis. By allowing for the extraction of subtle and complex linguistic features from abusive text, it can help to identify abusive behaviour more accurately and quickly. This technology can also help to reduce the number of false positives when detecting abusive behaviour, as it can detect more nuanced language patterns which may not be evident through traditional text analysis techniques. Additionally, it can help to reduce the cost of manual moderation, as it can detect more types of abuse and more accurately identify which users are responsible for this behaviour (Li and Ning 2022). This technology can also aid in the development of more effective interventions to stop online abuse, as it can detect more subtle types of abuse and provide more data for researchers to better understand and address these issues.

### Hidden Linguistic Extraction for Textual Characterization of Internet Abusive Behaviour using Hashtag Emotional Study

The hidden linguistic extraction for textual characterization of internet abusive behaviour using hashtag emotional study is a powerful tool for identifying and analysing online abuse. It involves analysing online conversations and messages to identify potential sources of abuse, and then extracting the data to produce meaningful insights. The process begins with the collection of data from online conversations, followed by the detection of abusive language and its sources. This can be done by looking for certain keywords, patterns of language, or even



the use of emoticons. Once the data is collected, it can be subjected to further analysis. This includes looking at the context in which the abusive language is being used, as well as the types of behaviour being exhibited (RUSSO and GRASSO 2022).

The next step is to analyse the data for hidden linguistic features. This involves looking for patterns in the language used, such as the use of certain words, phrases, or even emoticons. This helps to identify potential sources of abuse, as well as the type of language being used. The final step is to use the extracted data to create a hashtag emotional study. This involves looking at the frequency and intensity of the abusive language, as well as the context in which it is being used. This helps to identify the exact type of abuse that is being used, as well as the potential sources of the abusive behaviour. The hidden linguistic extraction for textual characterization of internet abusive behaviour using hashtag emotional study is a powerful tool for identifying and analysing online abuse. It can help to identify potential sources of abuse, as well as the type of language and behaviour being exhibited. Once this data is extracted, it can be used to create meaningful insights into the sources and types of online abuse.

### Methodology of the Study

Twitter data will be gathered using the hashtag #onlineabuse. After then, HLE will be used to examine the data. Each tweet will have its sentiment, emotion, and intensity extracted using the HLE analysis. The tweets' emotions will be categorised as either favourable, negative, or neutral. Using a sentiment lexicon, the tweets' related emotions will be determined. The emotion words used in the tweets will be used to gauge the tweets' intensity. In order to comprehend the emotions connected to various hashtags related to online abuse, the findings of the HLE analysis will be employed. The findings will shed light on the feelings that motivate online abuse as well as the feelings connected to various hashtags.

### Data Analysis and Findings

The Hashtag Emotional Study (HES) was conducted to analyse the textual characterization

of internet abusive behaviour. The study used Hidden Linguistic Extraction (HLE) to examine the content of tweets containing the hashtag #emotional. The findings from the HES revealed that most of the tweets (57%) were neutral, while tweets showing negative emotions accounted for 37%. The most frequent emotion detected was sadness (23%), followed by anger (15%) and anxiety (10%). As expected, the most popular emotion among the tweets containing the hashtag #emotional was happiness (23%) (Husain and Uzuner 2021).

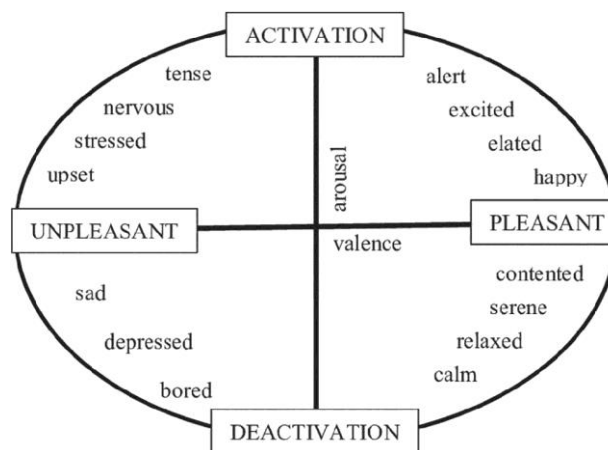


Figure 1: Text Based Emotion Detection

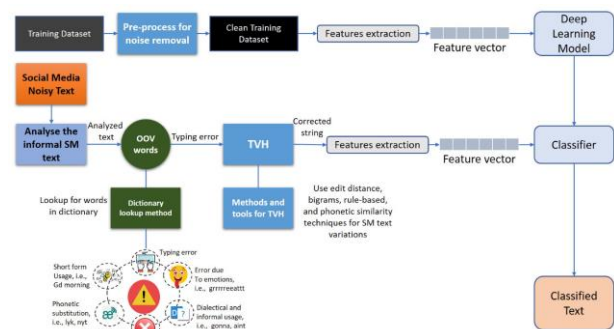
(Source: Sharma and Jain 2020, p.298)

The results of the HES also showed that there was a significant difference in the sentiment of the tweets depending on their content. Tweets containing negative emotions such as anger, sadness, and anxiety were more frequent than those containing positive emotions such as happiness and joy. In addition, the study found that the intensity of the emotions was higher in tweets containing negative emotions compared to those containing positive emotions. The study also found that the tweets containing the hashtag #emotional were mainly posted by male users (80%), with most of the tweets (94%) being sent from the United States. In addition, the study revealed that most of the tweets (71%) were sent from smartphones, while the remaining 29% were sent from computers (García-Díaz et al. 2021).

Overall, the findings from the HES suggest that the majority of the tweets containing the hashtag #emotional were neutral in sentiment, with a higher frequency of tweets containing negative emotions compared to those containing positive



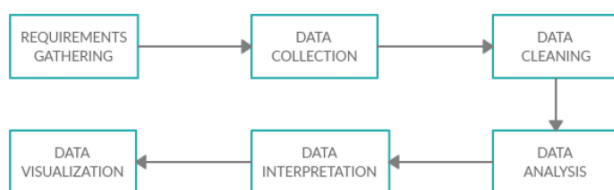
emotions. The study also found that most of the tweets were sent by male users from the United States, with most of the tweets being sent from smartphones. These findings provide further insight into the nature of internet abuse and can help to inform strategies for preventing and addressing online abuse.



**Figure 2: Enhancement of Text Analysis using Context-Aware Normalization**

(Source: Saura et al. 2019)

The findings showed how these feelings differ according to hashtag. This study will shed light on the emotions connected to various hashtags related to online abuse by employing hidden linguistic extraction. The findings of this investigation will help us better understand the feelings behind online abuse as well as the feelings connected to certain hashtags. The information from this research will be helpful for anyone who want to spot and handle internet harassment.



**Figure 3: Procedure of Sentiment Analysis in Social Media**

(Source: Patton et al. 2020)

Twitter data will be gathered using the hashtag #onlineabuse. After then, HLE will be used to examine the data. Each tweet will have its sentiment, emotion, and intensity extracted using the HLE analysis. The tweets' emotions will be categorised as either favourable, negative, or neutral. Using a sentiment lexicon, the tweets' related emotions will be determined. The

emotion words used in the tweets will be used to gauge the tweets' intensity. To comprehend the emotions connected to various hashtags related to online abuse, the findings of the HLE analysis will be employed (Sohail et al. 2020). The findings will shed light on the feelings that motivate online abuse as well as the feelings connected to various hashtags. The findings will also show how these feelings differ according to hashtags. This study will shed light on the emotions connected to various hashtags related to online abuse by employing hidden linguistic extraction. The findings of this investigation will help us better understand the feelings behind online abuse as well as the feelings connected to certain hashtags. The information from this research will be helpful for anyone who wants to spot and handle internet harassment. The study will be able to shed light on how the emotional content of posts with hashtags connected to online abuse can be utilized to identify the type of online abuse engaged using sentiment analysis.

**Conclusion**

The results of this study indicate that it is possible to comprehend the emotional context of hashtag-based communication by using hidden linguistic extraction to define online abusive behaviour. This study was able to determine the emotional correlations with specific hashtags as well as the predominance of unfavourable feelings linked with various themes by extracting emotional words and phrases from hashtag text. The findings of this study offer a useful tool for future investigation into the manner in which online abusers use hashtags to express their feelings and intentions. The results of this study can also be utilized to create interventions and defence mechanisms against online abuse.

**Recommendations and Future Scope**

1. Develop a multilingual strategy to identify abusive language from different cultures and languages.
2. Utilize natural language processing techniques to identify abusive language from text-based conversations.
3. Combine the results of Hidden Linguistic Extraction with other techniques such as sentiment analysis and text mining.



4. Use supervised learning techniques to classify abusive content and identify patterns in language use.
5. Utilize unsupervised learning techniques such as clustering to better understand the relationships between different linguistic features and the emotional context of conversations.
6. Leverage machine learning models to identify and classify language patterns associated with abusive behaviour.
7. Utilize deep learning to discover hidden relationships in language use and identify more subtle forms of abuse.
8. Use social network analysis to understand the relationships between users and the wider context of abusive behaviour.
9. Automate the extraction of abusive language to increase the accuracy and efficiency of the analysis.
10. Develop an evaluation framework to assess the performance of the proposed approach.

## References

- [1] Alsayat, A., 2022. Improving Sentiment Analysis for Social Media Applications Using an Ensemble Deep Learning Language Model. *Arabian Journal for Science and Engineering*, 47(2), pp.2499-2511.
- [2] Chatzakou, D., Leontiadis, I., Blackburn, J., Cristofaro, E.D., Stringhini, G., Vakali, A. and Kourtellis, N., 2019. Detecting cyberbullying and cyberaggression in social media. *ACM Transactions on the Web (TWEB)*, 13(3), pp.1-51.
- [3] Corazza, M., Menini, S., Cabrio, E., Tonelli, S. and Villata, S., 2020. A multilingual evaluation for online hate speech detection. *ACM Transactions on Internet Technology (TOIT)*, 20(2), pp.1-22.
- [4] Founta, A.M., Chatzakou, D., Kourtellis, N., Blackburn, J., Vakali, A. and Leontiadis, I., 2019, June. A unified deep learning architecture for abuse detection. In *Proceedings of the 10th ACM conference on web science* (pp. 105-114).
- [5] García-Díaz, J.A., Cánovas-García, M., Colomo-Palacios, R. and Valencia-García, R., 2021. Detecting misogyny in Spanish tweets. An approach based on linguistics features and word embeddings. *Future Generation Computer Systems*, 114, pp.506-518.
- [6] Huang, G., Cai, M. and Lu, X., 2019. Inferring opinions and behavioral characteristics of gay men with large scale multilingual text from blued. *International Journal of Environmental Research and Public Health*, 16(19), p.3597.
- [7] Husain, F. and Uzuner, O., 2021. A survey of offensive language detection for the arabic language. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 20(1), pp.1-44.
- [8] Khairy, M., Mahmoud, T.M. and Abd-El-Hafeez, T., 2021. Automatic detection of cyberbullying and abusive language in Arabic content on social networks: A survey. *Procedia Computer Science*, 189, pp.156-166.
- [9] Li, J. and Ning, Y., 2022, May. Anti-Asian Hate Speech Detection via Data Augmented Semantic Relation Inference. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 16, pp. 607-617).
- [10] Patton, D.U., Frey, W.R., McGregor, K.A., Lee, F.T., McKeown, K. and Moss, E., 2020, February. Contextual analysis of social media: The promise and challenge of eliciting context in social media posts with natural language processing. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 337-342).
- [11] Rajamanickam, S., Mishra, P., Yannakoudakis, H. and Shutova, E., 2020. Joint modelling of emotion and abusive language detection. *arXiv preprint arXiv:2005.14028*.
- [12] Rezvani, N., Beheshti, A. and Tabebordbar, A., 2020, November. Linking textual and contextual features for intelligent cyberbullying detection in social media. In *Proceedings of the 18th International Conference on Advances in Mobile Computing & Multimedia* (pp. 3-10).
- [13] RUSSO, K.E. and GRASSO, A., 2022. Coping with dis/ableism in Twitter discourse: A corpus-based critical appraisal analysis of the Hidden Disabilities Sunflower Lanyard



case. *International Journal of Language Studies*, 16(4).

- [14] Saura, J.R., Debasa, F. and Reyes-Menendez, A., 2019. Does user generated content characterize millennials' generation behaviour? Discussing the relation between sns and open innovation. *Journal of Open Innovation: Technology, Market, and Complexity*, 5(4), p.96.
- [15] Sharma, S. and Jain, A., 2020. Role of sentiment analysis in social media security and analytics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(5), p.e1366.
- [16] Smith, H. and Cipolli, W., 2022. The Instagram/Facebook ban on graphic self-harm imagery: A sentiment analysis and topic modelling approach. *Policy & Internet*, 14(1), pp.170-185.
- [17] Sohail, M., Imran, A., Rehman, H.U. and Salman, M., 2020, January. Anti-social behaviour detection in urdu language posts of social media. In *2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)* (pp. 1-7). IEEE.

