EmoMix-3L: A Code-Mixed Dataset for Bangla-English-Hindi Emotion Detection

Nishat Raihan*, **Dhiman Goswami***, **Antara Mahmud**,
Antonios Anastasopoulos, **Marcos Zampieri**

George Mason University, USA
{mraihan2, dgoswam}@gmu.edu

Abstract

Code-mixing is a well-studied linguistic phenomenon that occurs when two or more languages are mixed in text or speech. Several studies have been conducted on building datasets and performing downstream NLP tasks on code-mixed data. Although it is not uncommon to observe code-mixing of three or more languages, most available datasets in this domain contain code-mixed data from only two languages. In this paper, we introduce EmoMix-3L, a novel multi-label emotion detection dataset containing code-mixed data from three different languages. We experiment with several models on EmoMix-3L and we report that MuRIL outperforms other models on this dataset.

Keywords: Code Mixing, Dataset, Emotion Detection

1. Introduction

The ability to convey emotions is an essential part of human communication. NLP models have been applied to detect emotions (e.g., anger, fear, joy) in texts from social media (Gaind et al., 2019), customer service (Gupta et al., 2010), and healthcare (Ayata et al., 2020). Emotion detection is an important part of social media analysis and mining efforts that include popular tasks such as sentiment analysis (Liu, 2020) and stance detection (Kawintiranon and Singh, 2021).

Most studies on sentiment analysis and emotion detection are carried out in one language at a time (Abdul-Mageed and Ungar, 2017; Chatterjee et al., 2019). Apart from a few notable exceptions (Vedula et al., 2023), detecting emotion in multilingual and code-mixed environments has not been significantly explored. Code-mixing is very common in multilingual societies. It is defined as the practice of using words and grammatical constructions from two or more languages interchangeably (Muysken, 2000). Code-mixing can occur at various levels such as intra-sentential where code-mixing is present within a sentence, and inter-sententialwhere code-mixing is present across sentences.

Detecting sentiments and emotions in code-mixed texts is a challenging task that we address in this paper by introducing EmoMix-3L, a multi-label emotion detection containing Bangla, Hindi, and English code-mixed texts. These three languages are often used together by the population of West Bengal. They are also used by populations from South East Asian living in other parts of the world where English is spoken as the official language or *lingua franca* such London, New York, or Singapore. Recent studies have created resources for these three languages in tasks such as sentiment analysis and offensive language detection (Raihan et al., 2023a; Goswami et al., 2023). To the best of our knowledge, however, no datasets for emotion detection in Bangla-English-Hindi code-mix exists and EmoMix-3L fills this gap.

The main contributions of this paper are as follows:

- We introduce EmoMix-3L[1], a novel three-language code-mixed test dataset in Bangla-Hindi-English for multi-label emotion detection. EmoMix-3L contains 1,071 instances annotated by speakers of the three languages. We make EmoMix-3L freely available to the community.

- We provide a comprehensive evaluation of several monolingual, bilingual, and multilingual models on EmoMix-3L.

We present EmoMix-3L exclusively as a test set due to the unique and specialized nature of the task. The size of the dataset, while limited for training purposes, offers a high-quality testing benchmark with gold-standard labels. Given the scarcity of similar datasets and the challenges associated with data collection, EmoMix-3L provides an important resource for the evaluation of multi-label emotion detection models, filling a critical gap in multi-level code-mixing research.

[1] https://github.com/GoswamiDhiman/EmoMix-3L

WILDRE-7 Workshop, pages 11–16

2. Related Work

A few studies studies have addressed emotion detection on bilingual code-mixed data (Wadhawan and Aggarwal, 2021; Vedula et al., 2023; Ameer et al., 2022). Vedula et al. (2023) implemented a multi-class emotion detection model leveraging transformer-based multilingual Large Language Models (LLMs) for English-Urdu code-mixed text. However, the study's ability to interpret code-mixed sentences that combine English and Roman Urdu had limitations. The study by Ameer et al. (2022) highlights how multi-label emotion classification may be used to identify every emotion that could exist in a given text. 11,914 code-mixed (English and Roman Urdu) SMS messages make up the substantial benchmark corpus presented in this paper for the multi-label emotion classification challenge.

There have been a number studies on Bengali-English code-mixed data. Mursalin et al. (2022) used deep learning approaches to identifying emotions from texts containing mixed Bengali and English coding, with an emphasis on comparing and contrasting the effectiveness of the suggested model with other ML and DL methods already in use. Ahmad et al. (2019) have explored and analyzed regional Indian code-mixed data. In this paper, the importance and applications of sentiment detection in a variety of domains are discussed, with an emphasis on Indo-Aryan languages like Tamil, Bengali, and Hindi.

A few studies have addressing Bengali-English-Hindi code-mixing on social media. Raihan et al. (2023a) uses multiple monolingual, bilingual, and multilingual models and a unique dataset with gold standard labels for sentiment analysis in Bangla-English-Hindi. Goswami et al. (2023) presents a novel offensive language identification dataset with the same three languages. Finally, another similar work include the TB-OLID dataset (Raihan et al., 2023b) that contains both transliterated and code-mixed data for offensive language identification.

3. The *EmoMix-3L* Dataset

We choose a controlled data collection method, asking the volunteers to freely contribute data in Bangla, English, and Hindi. This decision stems from several challenges of extracting such specific code-mixed data from social media and other online platforms. Our approach ensures data quality and sidesteps the ethical concerns associated with using publicly available online data. Such types of datasets are often used when it is difficult to mine them from existing corpora. As examples, for fine-tuning LLMs on instructions and conversations, semi-natural datasets like Databricks (2023) and Nie (2023) have become popular.

Data Collection Ten undergraduate students fluent in the three languages was asked to prepare 300 to 350 social media posts each. They were allowed to use any language, including Bangla, English, and Hindi to prepare posts on several daily topics like politics, sports, education, social media rumors, etc. We also ask them to switch languages if and wherever they feel comfortable doing so. The inclusion of emojis, hashtags, and transliteration was also encouraged. The students had the flexibility to prepare the data as naturally as possible. Upon completion of this stage, we gathered 2,598 samples that contained at least one word or sub-word from each of the three languages using langdetect (Mazzocchi, 2012) an open-sourced Python tool for language identification.

Data Annotation We annotate the dataset in two steps. Firstly, we recruited three students from social science, computer science, and linguistics fluent in the three languages to serve as annotators. They annotated all 2,598 samples with one of the five labels (Happy, Surprise, Neutral, Sad, Angry) with a raw agreement of 47.9%. We then take 1,246 instances, where all three annotators agree on the labels, and use them in a second step. To further ensure high-quality annotation, we recruit a second group of annotators consisting of two NLP researchers fluent in the three languages. After their annotation, we calculate a raw agreement of 86% (Kvålseth, 1989), a Cohen Kappa score of 0.72. After the two stages, we only keep the instances where both annotators agree, and we end up with a total of 1,071 instances. The label distribution is shown in Table 1.

Label	Instances	Percentage
Happy	228	21.29%
Surprise	227	21.20%
Neutral	223	20.82%
Sad	205	19.14%
Angry	188	17.55%
Total	1,071	100%

Table 1: Label distribution in EmoMix-3L

Dataset Statistics A detailed description of the dataset statistics is provided in Table 2. Since the dataset was generated by people whose first language is Bangla, we observe that the majority of tokens in the dataset are in Bangla. There are several *Other* tokens in the dataset that are not from Bangla, English, or Hindi language. The *Other* tokens in the dataset primarily contain transliterated words as well as emojis and hashtags. Also, there are several misspelled words that have been classified as *Other* tokens too.

	All	Bangla	English	Hindi	Other
Tokens	98,011	36,784	6,587	15,560	39,080
Types	21,766	9,118	1,237	1,523	10,022
Avg	91.51	34.35	6.15	14.53	36.49
Std Dev	20.24	9.13	2.88	5.94	10.64

Table 2: EmoMix-3L Data Card. The row *Avg* represents the average number of tokens with its standard deviation in row *Std Dev*.

Synthetic Train and Development Set We present EmoMix-3L as a test dataset and we build a synthetic train and development set that contains Code-mixing for Bangla, English, and Hindi. We use an English training dataset annotated with the same labels as EmoMix-3L, namely Social Media Emotion Dataset (SMED)[2]. We then use the *Random Code-mixing Algorithm* (Krishnan et al., 2022) and *r-CM* (Santy et al., 2021) to generate the synthetic Code-mixed dataset. Similar approach is also found in (Gautam et al., 2021).

4. Experiments

Monolingual Models We use six monolingual models for these experiments, five general models, and one task fined-tuned model. The five monolingual models are DistilBERT (Sanh et al., 2019), BERT (Devlin et al., 2019), BanglaBERT (Kowsher et al., 2022), roBERTa (Liu et al., 2019), HindiBERT (Nick Doiron, 2023). BanglaBERT is trained in only Bangla and HindiBERT in only Hindi while DistilBERT, BERT, and roBERTa are trained in English only. Finally, the English task fine-tuned model we use is emoBERTa (Kim and Vossen, 2021).

Bilingual Models BanglishBERT (Bhattacharjee et al., 2022) and HingBERT (Nayak and Joshi, 2022) are used as bilingual models as they are trained on both Bangla-English and Hindi-English respectively.

Multilingual Models We use mBERT (Devlin et al., 2019) and XLM-R (Conneau et al., 2020) as multilingual models which are respectively trained on 104 and 100 languages including Bangla-English-Hindi. We also use IndicBERT (Kakwani et al., 2020) and MuRIL (Khanuja et al., 2021) which cover 12 and 17 Indian languages, respectively, including Bangla-English-Hindi. We also perform hyper-parameter tuning while using all the models to prevent overfitting.

Prompting We use prompting with GPT-3.5-turbo model (OpenAI, 2023) from OpenAI for this task. We use the API for zero-shot prompting (see Figure 1) and ask the model to label the test set.

Additionally, we run the same experiments separately on synthetic and natural datasets splitting both in a 60-20-20 way for training, evaluating, and testing purposes.

Role: "You are a helpful AI assistant. You are given the task of Emotion Detection.

Definition: An emotion is a feeling that can be caused by the situation that people are in or the people they are with. You will be given a text to label either 'Happy', 'Surprise', 'Neutral', 'Sad' or 'Angry'.

Task: Generate the label for this **"text"** in the following format: <*label*> *Your_Predicted_Label* <*label*>. Thanks."

Figure 1: Sample GPT-3.5 prompt.

5. Results

In this experiment, synthetic data is used as a training set, and natural data is used as the test set. The F1 scores of monolingual models range from 0.14 to 0.41, where roBERTa performs the best. MuRIL is the best of all the multilingual models, with an F1 score of 0.54. Besides, a zero-shot prompting technique on GPT 3.5 turbo provides a 0.51 weighted F1 score. The task fine-tuned model emoBERTa provides the F1 score of 0.42. BanglishBERT scores 0.44 which is the best F1 score among all the bilingual models. These results are available in Table 3.

Models	F1 Score
MuRIL	**0.54**
XLM-R	0.51
GPT 3.5 Turbo	0.51
BanglishBERT	0.44
HingBERT	0.43
emoBERTa	0.42
roBERTa	0.41
BERT	0.38
mBERT	0.35
DistilBERT	0.24
IndicBERT	0.22
BanglaBERT	0.16
HindiBERT	0.14

Table 3: Weighted F-1 score for different models: training on synthetic and tested on natural data (EmoMix-3L).

We perform the same experiment using synthetic data for training and testing. We present results in Table 4. Here, MuRIL with 0.67 F1 score is the

[2] https://www.kaggle.com/datasets/gangulyamrita/social-media-emotion-dataset

best-performing model. BERT is the best among the monolingual models where their F1 range from 0.32 to 0.45. BanglishBERT with 0.47 F1 score is the best among the bilingual models. The task fine-tuned model emoBERTa scores 0.41 for the synthetic dataset.

Models	Weighted F1 Score
MuRIL	**0.67**
XLM-R	0.51
mBERT	0.49
BanglishBERT	0.47
HingBERT	0.45
BERT	0.44
emoBERTa	0.41
roBERTa	0.41
DistilBERT	0.40
BanglaBERT	0.39
IndicBERT	0.38
HindiBERT	0.32

Table 4: Weighted F-1 score for different models: training on synthetic and tested on synthetic data.

5.1. Error Analysis

The confusion matrix for the best-performing model MuRIL for training on synthetic and tested in EmoMix-3L is shown in Figure 2.

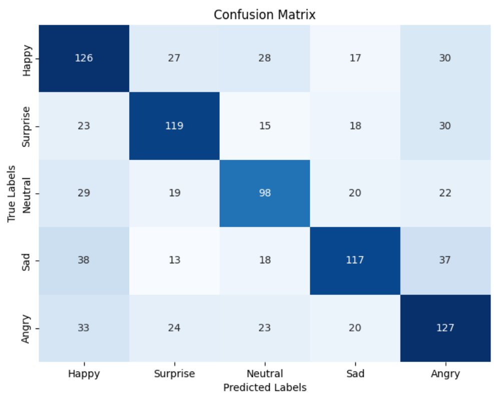

Figure 2: Confusion Matrix (Training on synthetic data, tested on EmoMix-3L).

We observe *Other* tokens in more than 39% of the whole dataset, as shown in Table 2. These tokens occur due to transliteration which poses a challenge for most of the models since not all of the models are pre-trained on transliterated tokens. BanglishBERT did well since it recognizes both Bangla and English tokens. However, the total number of tokens for Hindi-English is less than Bangla-English tokens, justifying HingBERT's inferior performance compared to BanglishBERT (see Table 3). Also, misspelled words and typos are also observed in the datasets, which are, for the most part, unknown tokens for the models, making the task even more difficult. Some examples are available in Appendix A which are classified wrongly by all the models.

6. Conclusion and Future Work

We introduce EmoMix-3L, a novel dataset containing 1,071 instances of Bangla-English-Hindi code-mixed content. We have also created 100,000 instances of synthetic data in the same languages to facilitate our training methods. We have tested multiple monolingual models on these datasets, and MuRIL performs the best, especially when it was trained on synthetic data and tested on EmoMix-3L. MuRIL was also the best in the scenario of both training and testing on synthetic data, outperforming all the other models in multi-label emotion detection. Looking ahead, we would like turning EmoMix-3L into a larger dataset serving both as a training and testing dataset. We would also like to create pre-trained tri-lingual code-mixing models. It will facilitate the emotion detection task in the intricate mix of Bangla, English, and Hindi. Moreover, we would like to explore the performance of large language models by fine-tuning them on code-mixed datasets. This will provide valuable insights into their unexplored training corpora and their ability to cope with code-mixed scenarios.

References

Muhammad Abdul-Mageed and Lyle Ungar. 2017. Emonet: Fine-grained emotion detection with gated recurrent neural networks. In *Proceedings of ACL*.

Gazi Imtiyaz Ahmad, Jimmy Singla, and Nikita Nikita. 2019. Review on sentiment analysis of indian languages with a special focus on code mixed indian languages. In *Proceedings of ICACTM*.

Iqra Ameer, Grigori Sidorov, Helena Gomez-Adorno, and Rao Muhammad Adeel Nawab. 2022. Multi-label emotion classification on code-mixed text: Data and methods. *IEEE Access*, 10:8779–8789.

Değer Ayata, Yusuf Yaslan, and Mustafa E Kamasak. 2020. Emotion recognition from multimodal physiological signals for emotion aware healthcare systems. *Journal of Medical and Biological Engineering*, 40:149–157.

Abhik Bhattacharjee, Tahmid Hasan, Wasi Ahmad, Kazi Samin Mubasshir, Md Saiful Islam, Anindya

Iqbal, M. Sohel Rahman, and Rifat Shahriyar. 2022. BanglaBERT: Language model pretraining and benchmarks for low-resource language understanding evaluation in Bangla. In *Findings of the ACL*.

Ankush Chatterjee, Kedhar Nath Narahari, Meghana Joshi, and Puneet Agrawal. 2019. Semeval-2019 task 3: Emocontext contextual emotion detection in text. In *Proceedings of SemEval*.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of ACL*.

Databricks. 2023. Dolly 2.0: An open source, instruction-following large language model. Accessed: 2023-09-10.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL*.

Bharat Gaind, Varun Syal, and Sneha Padgalwar. 2019. Emotion detection and analysis on social media. *arXiv preprint arXiv:1901.08458*.

Devansh Gautam, Prashant Kodali, Kshitij Gupta, Anmol Goel, Manish Shrivastava, and Ponnurangam Kumaraguru. 2021. Comet: Towards code-mixed translation using parallel monolingual sentences. In *Proceedings of CALCS*.

Dhiman Goswami, Md Nishat Raihan, Antara Mahmud, Antonios Anastasopoulos, and Marcos Zampieri. 2023. OffMix-3L: A novel code-mixed test dataset in bangla-english-hindi for offensive language identification. In *Proceedings of SocialNLP*.

Narendra Gupta, Mazin Gilbert, and Giuseppe Di Fabbrizio. 2010. Emotion detection in email customer care. In *Proceedings of NAACL*.

Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, NC Gokul, Avik Bhattacharyya, Mitesh M Khapra, and Pratyush Kumar. 2020. Indicnlpsuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for indian languages. In *Findings of the ACL*.

Kornraphop Kawintiranon and Lisa Singh. 2021. Knowledge enhanced masked language model for stance detection. In *Proceedings of NAACL*.

Simran Khanuja, Diksha Bansal, Sarvesh Mehtani, Savya Khosla, Atreyee Dey, Balaji Gopalan, Dilip Kumar Margam, Pooja Aggarwal, Rajiv Teja Nagipogu, Shachi Dave, et al. 2021. Muril: Multilingual representations for indian languages. *arXiv preprint arXiv:2103.10730*.

Taewoon Kim and Piek Vossen. 2021. Emoberta: Speaker-aware emotion recognition in conversation with roberta. *arXiv preprint arXiv:2108.12009*.

M Kowsher, Abdullah As Sami, Nusrat Jahan Prottasha, Mohammad Shamsul Arefin, Pranab Kumar Dhar, and Takeshi Koshiba. 2022. Bangla-bert: transformer-based efficient model for transfer learning and language understanding. *IEEE Access*, 10:91855–91870.

Jitin Krishnan, Antonios Anastasopoulos, Hemant Purohit, and Huzefa Rangwala. 2022. Cross-lingual text classification of transliterated hindi and malayalam. In *Proceedings of Big Data*.

Tarald O Kvålseth. 1989. Note on cohen's kappa. *Psychological reports*, 65(1):223–226.

Bing Liu. 2020. *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge university press.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Daniele Mazzocchi. 2012. langdetect: Language detection library. Python library.

Golam Sarwar Md Mursalin, Suborno Deb Bappon, and Muhammad Ibrahim Khan. 2022. A deep learning approach for recognizing textual emotion from bengali-english code-mixed data. In *Proceedings of ICCIT*.

Pieter Muysken. 2000. The study of code-mixing. *Bilingual Speech: A Typology of Code-Mixing*, 110.

Ravindra Nayak and Raviraj Joshi. 2022. L3Cube-HingCorpus and HingBERT: A code mixed Hindi-English dataset and BERT language models. In *Proceedings of WILDRE*.

Nick Doiron. 2023. hindi-bert. Accessed: 2023-09-10.

Jianzhi Nie. 2023. Awesome instruction datasets. Accessed: 2023-09-10.

OpenAI. 2023. Gpt-3.5 turbo fine-tuning and api updates. Accessed: 2023-08-28.

Md Nishat Raihan, Dhiman Goswami, Antara Mahmud, Antonios Anastasopoulos, and Marcos Zampieri. 2023a. SentMix-3L: A novel code-mixed test dataset in bangla-english-hindi for sentiment analysis. In *Proceedings of SEALP*.

Md Nishat Raihan, Umma Tanmoy, Anika Binte Islam, Kai North, Tharindu Ranasinghe, Antonios Anastasopoulos, and Marcos Zampieri. 2023b. Offensive language identification in transliterated and code-mixed bangla. In *Proceedings of BLP*.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of BERT: smaller, faster, cheaper and lighter. In *Proceedings of EMC2*.

Sebastin Santy, Anirudh Srinivasan, and Monojit Choudhury. 2021. Bertologicomix: How does code-mixing interact with multilingual bert? In *Proceedings of AdaptNLP*.

Bhaskara Hanuma Vedula, Prashant Kodali, Manish Shrivastava, and Ponnurangam Kumaraguru. 2023. Precogiiith@ wassa2023: Emotion detection for urdu-english code-mixed text. In *Proceedings of WASSA*.

Anshul Wadhawan and Akshita Aggarwal. 2021. Towards emotion recognition in hindi-english code-mixed data: A transformer based approach. *arXiv preprint arXiv:2102.09943*.

A. Examples of Misclassified Instances

Happy: Finally got the চাবি to our new বাড়ি! So excited to start making স্মৃতি in our new space. #HomeSweetHome #NewBeginnings मैं उस अद्भुत सपोर्ट सिस्टम के लिए आभारी हूं जिसने इस यात्रा के माध्यम से मेरी मदद की। हमेशा मेरे लिए रहने के लिए धन्यवाद। #आभारी #धन्य আই এম থ্রিলড তাে এনাউন্স দ্যাট আই হ্যাভ অফিসিয়ালি কমপ্লেটেড মাই মেডিটেশন চ্যালেঞ্জ! ফিলিং মারে সেন্টারেড এন্ড গ্রাউন্ডেড দ্যান এভার বছরের পর বছর saving and budgeting করার পর, আমি ঘাষেণা করতে পেরে রামোঞ্চিত যে আমি আমার ছাত্র ঋণ পরিশাধে করেছি! #Debtfree #FinancialFreedom main kee apanee haal kee yaatra par kee gaee avishvasaneey yaadon ke lie bahut aabhaaree hoon. vaapas jaane ke lie intajaar nahin kar sakata! #travailgoals #advainturai

Suprise: একটি শান্তিপূর্ণ সমুদ্র সৈকতে হাঁটা, আপনার পায়ের আঙ্গুলের মধ্যে বালি অনুভব করা এবং ঢেউয়ের শব্দর মধ্যে something magical ফিল্মন অর পারিয়ন কী কাহানিয়ন কী সামগ্রী হয, লেকিন বাস্তবিক জীবন মেইন বহি সাকাতী হাই सोच में खोए हुए समुद्र तट पर चलने की कल्पना करें, जब कुछ आपकी नज़र में आ जाए। Hidden treasures are waiting to be discovered, যদি আমাদের চাখে থাকে তাদের দেখার ajke somudror tire akta sundor hater kacher kaner dul peyechi

Neutral: ইমেল চেক while riding the tram to the office. व्यस्त ट्रेन स्टेशन पर भीड़ के माध्यम से नेविगेट करना। ওয়াচিং এ মুভি ও আ ট্যাবলেট ডুরিং টি বাস রাইড টু ওয়ার্ক। ব্যস্ত train স্টেশনে ভিড়ের মধ্য দিয়ে নেভিগেট করা। paark kee bench par baithakar bas ke aane ka intajaar kar rahe the.

Sad: একটি সম্পর্কের সমাপ্তি এবং ভালবাসা হারানারে মত situation মেনে নেয়ার মত না It makes you lonely আপকে জিবান মে হাজার লগ হনেকি বাদবি আপ ওহি এক ইনসান ক ইয়াদ কারংগি হার ওয়াক্ত নুকসান কা শোক मनाना और ठीक होने के लिए आवश्यक समय लेना ठीक है। Memories flood our minds, and we find ourselves yearning for something যা আমরা আর পাব না। jake valobaschi se amar sathe sob somoy thakbe na aita kokhon o vabini

Angry: আই এম এংরি রাইট নাউ বিকজ ই জাস্ট রেসিভড আ প্যাসিভ-এগ্রেসিভ কমেন্ট ফ্রম সামওয়ান! এটা অবিশাস্য যে কিভাবে some man how অভদ্র এবং অসম্মানজনক হতে পারে। I don't deserve to be আচরণ করার যাগ্যে way and I won't stand for it. यदि आपको मुझसे कोई समस्या है, तो मेरी पीठ पीछे भद्दी टिप्पणि किरने के बजाय इसे सीधे संबोधित करने की शालीनता रखें। main kisee aur kee nakaaraatmakata ko mujhe neeche laane se mana karata hoon, lekin gambheerata se, bada hokar seekhata hoon ki ek paripakv vayask kee tarah kaise sanvaad karana hai.