BanglaBait: Semi-Supervised Adversarial Approach for Clickbait Detection on Bangla Clickbait Dataset

Md. Motahar Mahtab
BRAC University
Dhaka, Bangladesh
mahtab27672767@gmail.com

Monirul Haque
BRAC University
Dhaka, Bangladesh
monirul.haque.mail@gmail.com

Mehedi Hasan
BRAC University
Dhaka, Bangladesh
mehedi.hasan@g.bracu.ac.bd

Farig Sadeque
BRAC University
Dhaka, Bangladesh
farig.sadeque@bracu.ac.bd

Abstract

Intentionally luring readers to click on a particular content by exploiting their curiosity defines a title as clickbait. Although several studies focused on detecting clickbait titles in English articles, low-resource language like Bangla has not been given adequate attention. To tackle clickbait titles in Bangla, we have constructed the first Bangla clickbait detection dataset containing 15,056 labeled news articles and 65,406 unlabelled news articles extracted from clickbait-dense news sites. Each article has been labeled by three expert linguists and includes an article's title, body, and other metadata. By incorporating labeled and unlabelled data, we finetune a pre-trained Bangla transformer model in an adversarial fashion using Semi-Supervised Generative Adversarial Networks (SS-GANs). The proposed model acts as a good baseline for this dataset, outperforming traditional neural network models (LSTM, GRU, CNN) and linguistic feature-based models. We expect that this dataset and the detailed analysis and comparison of these clickbait detection models will provide a fundamental basis for future research into detecting clickbait titles in Bengali articles. We have released the corresponding code and dataset [1].

1 Introduction

Due to the widespread usage of the internet, the news industry has progressively evolved into an online news industry leading to the explosion of clickbait titles in recent years. As the concept of clickbait can be hazy to grasp, the classification of clickbait is a highly subjective endeavor. Biyani et al. (2016) suggests that clickbait titles can be roughly categorized into eight types. Table 1 displays this different clickbait categories[2] and their corresponding Bangla articles.

There are an estimated 11.4 million internet users in Bangladesh[3] who receive their daily news mostly from online news sites. However, no research has been conducted on tackling the increasing number of clickbait titles on these sites and other news websites. For English articles, Potthast et al. (2018a) built the first large-scale annotated clickbait corpus (Webis Clickbait Corpus 2017) containing 338,517 articles. In the Bangla language, the lack of an annotated clickbait-rich dataset is hindering the progress of Bangla Clickbait Detection. We construct the first Bangla Clickbait Corpus, which contains an article's title, content, and other metadata collected from various clickbait-rich websites upon which future researchers can build an effective Bangla clickbait detection model. The effectiveness of Semi-Supervised Generative Adversarial networks (SS-GANs; Salimans et al., 2016) have been shown for text classification tasks in Croce et al. (2020). From our experiments, it is evident that fine-tuning a Bangla ELECTRA model in this setup improves clickbait detection performance outperforming all other model types.

The main contributions of this paper can be summarized as follows:

[1] https://github.com/mdmotaharmahtab/BanglaBait

[2] 'wrong' category in Biyani et al. (2016) was replaced by 'question' category - reason described in details in section 3.1

[3] https://www.cia.gov/the-world-factbook/countries/bangladesh/

Proceedings of Recent Advances in Natural Language Processing, pages 748–758
Varna, Sep 4–6, 2023
https://doi.org/10.26615/978-954-452-092-2_081

Category	Reason	Headline example & Translation
Questions	Titles pose a query that compels the reader to click to get the answer.	হঠাৎ উধাও সালমানের নায়িকা রম্ভা, কী করছেন তিনি? (Salman's actress Rombha mysteriously disappeared, what is she up to?).
Inflammatory	Titles evoke strong emotion.	সপাটে লাথি মেরে স্ট্যাম্প ভেঙ্গে আম্পায়ারকে গালি! (Lost his cool, kicked, then busted out the stamps before abusing the umpire!)
Curiosity Gap/Teasing	Titles leave the reader in the dark, which tempts them to click.	জেনে নিন ফ্রিজ ছাড়াই দীর্ঘদিন মাংস সংরক্ষণের উপায়! (Explore how to preserve meat without a refrigerator!)
Ambiguous	Imprecise or unclear titles that pique interest.	মীরও ছাড় দিলেন না নুসরাতকে (Not even Mir spared Nusrat)
Exaggerate	Titles overstating what is written on the landing page.	জামের সঙ্গে যে তিন খাবার খেলে হতে পারে মৃত্যুও! (Three foods when combined with blackberries, could kill you!)
Graphic	Salacious, unsettling, or implausible subject matter.	ছেলের হাতে পরকীয়ায় ধরা পরায় মা ছেলেকে কেটে বস্তায় ভরে পানিতে ফেলে দেয় (After he finds her cheating, the mother cuts her son into bits and stuffs him into a bag before tossing it into the water.)
Formatting	Excessive use of punctuation or other symbols.	ফারিয়া 'আউট' পরীমনি 'ইন'! (Faria 'out' Porimoni 'in'!)
Bait & Switch	Overpromising titles with under-delivering content; requires additional clicks.	এক শরীরে দুই প্রাণ! একজন ইংরেজি শিক্ষক অপরজন গণিতের (One body two souls! One is an English teacher, whereas another is of Mathematics.)

Table 1: Clickbait news titles and their categories.

- We create an annotated dataset of 15,056 articles and an unannotated dataset of 65,406 Bangla articles rich with clickbait titles. The dataset contains the title, body, domain, article category, publication date, and English translation of title and content. We plan to release both of these datasets upon acceptance of the paper.
- We develop the first Bangla Clickbait Detection model for Bangla textual data by thoroughly experimenting with different statistical machine learning algorithms, deep neural networks using state-of-the-art embeddings, and Transformer networks (Vaswani et al., 2017) to discover the best approach for detecting clickbait. Section 7 analyzes the quantitative comparisons among all these different models.
- We train a Bangla Transformer model in a Semi-Supervised Generative Adversarial setup and show that it improves upon existing models trained in a supervised manner.

2 Related Work

The origin of clickbait is rooted in tabloids which have been in journalism since the 1980's (Bird, 2008). Generally, clickbait detection features can be obtained from 3 different origins: clickbait teaser phrase or post text, the attached article that the post text wants the user to click, and metadata for both (Potthast et al., 2018a). Apart from the post text, which is used by most to identify clickbait, the works of Potthast et al. (2016) and Biyani et al. (2016) also considered the linked article, metadata and used handcrafted features, TF-IDF similarity between headline and article content and Gradient Boosted Decision Trees (GBDT). Potthast et al. (2018a) suggested that clickbait detection should be a regression problem instead of a binary classification challenge, as the latter provides a way to measure how much clickbait is in the teaser message. They initiated the Webis clickbait challenge 2017, which boosted research activity in clickbait detection giving rise to highly effective and flexible deep learning techniques. For clickbait challenge 2017, Zhou (2017) first used self-attentive RNN (Elman, 1990) to select the important words in the title and created a BiGRU (Cho et al., 2014) network to encode the contextual information. Thomas (2017), on the other hand, incorporated article content into an LSTM model (Hochreiter and Schmidhuber, 1997) for the clickbait challenge. Rony et al. (2017) used continuous skip-gram model (Mikolov et al., 2013) to generate the word embedding of clickbait titles. However, Indurthi et al. (2020) first investigated the application of transformer regression models in clickbait detection and achieved the first position in the clickbait challenge. Besides, Hossain

et al. (2020) created the first Bengali newspaper dataset for Bengali fake news detection containing an annotated dataset of $\approx 50K$ Bangla news. To the best of our knowledge, the first attempt to detect clickbait in Bangla was made by Munna and Hossen (2021). They created a dataset on video-sharing platforms containing Bangla and English video links and used numerical features to detect clickbait links. However, no research has been conducted to tackle clickbaits in written news mediums using the textual features of the article. We present the first clickbait detection dataset in Bangla and also provide a comprehensive comparison of various models to detect them.

3 First Bangla Dataset for Detecting Bangla Clickbait News Articles

3.1 Data Collection

We first compile a list of websites that publish Bangla news articles. Although Potthast et al. (2018b, 2016) used metrics like the number of retweets to select the most influential websites, such metric providing services like Alexa ranking[4] is unavailable for most prominent Bangla Websites. Instead, we first create a preliminary list of Bangla news article sites from where we choose a website for scraping if the homepage seems to contain more clickbait than non-clickbait titles after a cursory glance by the annotators. We also select some famous Bengali online news publishers such as Kaler Kantha[5], SomoyTV[6], and RTV news[7] for scraping to facilitate future investigation into clickbait practices in popular Bangla news mediums. Before scraping, we check whether the publishers we select have terms and conditions against scraping or using their content for educational or research purposes to avoid copyright infringement. Utilizing the Python Selenium module, we have scraped data from the first week of February 2019 to the last week of February 2022.

Although Hossain et al. (2020) published the first dataset of Bangla Fake news, we find it necessary to create a separate dataset for clickbait in Bangla as a news title can be a clickbait without necessarily being fake news (Dong et al., 2019) [8]. To enrich our dataset size, one thousand titles labeled 'clickbait' from Bangla Fake News Dataset (Hossain et al., 2020) are added to our own dataset after their labels are revised again by annotators.

3.2 Annotation Process

The dataset is annotated by three annotators with an MA in Bangla Linguistics. At first, they study the annotations of popular English clickbait datasets (Potthast et al., 2018b; Agrawal, 2016; Potthast et al., 2016). Investigating English titles help the annotators understand how titles induce curiosity in practice, which they can then use to annotate Bangla titles. As questions naturally entice interest, a new clickbait category named 'question' is added to the clickbait categories in Table 1. No publisher or source of the article is available to the annotators to avoid any induced publisher-based biases as reported by Potthast et al. (2018a) to be the case for several clickbait datasets (Rony et al., 2017; Ganguly, 2016; Agrawal, 2016). A majority vote among the annotators decides the final annotation. The annotators reach an inter-annotator agreement Fleiss kappa (Fleiss et al., 1971) of 0.62, which is substantial Landis and Koch (1977) and enough for a good speculative conclusion regarding annotator agreement (Artstein and Poesio, 2008).

The annotators mark clickbait news as a numeric value of 1 and non-clickbait news as a numeric value of 0. Our labeled and unlabelled datasets contain eight categories - Economy, Education, Entertainment, Politics, International, Sports, National, and Science & Technology of clickbait and non-clickbait titles. After removing all duplicates from labeled and unlabeled datasets, our dataset contains 15,056 unique news articles with 9,817 non-clickbait and 5,239 clickbait articles, and 65,406 unique unlabelled articles. The labeled and unlabeled datasets do not have any overlapping content or titles. The test set is further curated by removing titles that have similar titles in the training set through Levenshtein distance (Levenshtein, 1965). Table 2 shows that clickbait titles have a slightly higher average number of words and punctuation than non-clickbait titles. The most frequent fifteen words in clickbait titles are -

এই (this), যে (that), না (no), ভাইরাল (viral),
ভিডিও (video), যা (which), করে (does), নিয়ে (with),
বিয়ে (marriage), থেকে (from), সেই (that), এক (one),
তুমুল (intense), কি (what), করতে (do)

It contains words like viral, video, and intense which usually induce readers to click. Each data instance contains the title and content of the article, publishing date, domain, news category, translated

[4] https://www.alexa.com/
[5] https://www.kalerkantho.com/
[6] https://www.somoynews.tv/
[7] https://www.rtvonline.com/
[8] More details can be found in section A.3

Information	Value	
Crawling Period	Feb 2019 - Feb 2022	
Total Clickbait	5239	
Total Non-clickbait	9817	
Total Unlabelled	65406	
Title Analysis	**Clickbait**	**Non-clickbait**
Average number of characters	52.845	49.097
Average number of words	8.983	7.8356
Average word length	4.99	5.4
Average Punctuation	1.003	0.805

Table 2: Summary statistics of our dataset.

Column	Value
Domain	https://www.rtvonline.com/
Date	2021-05-25
Title	মাত্র ১৩ টাকায় মিলছে বাড়ি!
Content	শুনতে অবাক লাগলেও এটাই সত্য। মাত্র ১৩ টাকায় কেনা যাবে বাড়ি। ফুটবলের সুবাদে অনেকেই ক্রোয়েশিয়ার নাম জানেন। সেই দেশেই মাত্র ১৩ টাকায় কেনা যাবে বাড়ি। দেশটির লেগ্রাড শহর এমন অবিশ্বাস্য অফার দিয়েছে। খবর হিন্দুস্তান টাইমসের।...
Label	1 (Clickbait)
Translated Title	It's only Rs. 13!
Translated Content	That's the truth, though it sounds surprising. Only 13 rupees can be bought at home...
Category	Science & Technology

Table 3: Sample Data

title, and translated content as shown in Table 3.

4 Human Baseline

Five human annotators who are undergraduate and regular newspaper readers are given 200 news article titles from the test set to annotate. They achieve an inter-annotator agreement of Fleiss' kappa (Fleiss et al., 1971) score of 0.374, which is fair according to Landis and Koch (1977). Compared to our dataset annotators' score, this score is much lower. Our annotation process includes investigating English titles first to better form a coherent perception of clickbait titles. By majority voting among the five annotators, we select the final labels and achieve an F1 score of 76.82% and an accuracy of 77.01% on the clickbait class, which serves as the human baseline for Bangla clickbait detection shown in Table 4.

5 Approach

5.1 GAN-BanglaBERT

In Generative Adversarial Network (Goodfellow et al., 2014), a generator $\mathcal{G}$ is trained to generate a data distribution similar to the real data to 'fool' the discriminator $\mathcal{D}$ and $\mathcal{D}$ is trained to differentiate between the two in an adversarial fashion. Semi-Supervised GANs (SS-GANs; Salimans et al., 2016) train the discriminator $\mathcal{D}$ to predict the classification labels along with the additional task of predicting whether the data is real or fake. This training technique helps the model improve its inner representations by utilizing the unlabelled and generated data (Croce et al., 2020). Following researchers of Croce et al. (2020), we finetune a BanglaBERT (Bhattacharjee et al., 2021), a state-of-the-art ELECTRA (Clark et al., 2020) model pre-trained on 35 GB of Bangla textual data from the web and call it 'GAN-BanglaBERT' throughout the paper. Figure 1 shows the overall architecture of the GAN-BanglaBERT model. Generator $\mathcal{G}$ and discriminator $\mathcal{D}$ both are a 2-layered deep neural network(DNN). A 100-dimensional noise vector is drawn from a standard normal distribution $N\left(\mu = 0, \sigma^2 = 1\right)$ following the initialization practice in GANs (Goodfellow et al., 2014). Generator $\mathcal{G}$ produces $h_{fake} \in R^d$ vector from this noise vector where d is the last layer size of the pre-trained Transformer network. Discriminator $\mathcal{D}$ takes in input the concatenation of both real and fake data's representation $[h_{real}; h_{fake}]$. Detailed training loss calculation is provided in Croce et al. (2020), which remains unchanged in our implementation. The average of the last hidden layer outputs of BanglaBERT is the transformer encoding h_{real} for a real title.

5.2 Comparison Methods

We compare the GAN-BanglaBERT model to the following models.

- Statistical Models: For statistical methods, we employ a Logistic and Random Forest classifier on a combination of various features like TF-IDF (term frequency–inverse document frequency) of the word and character n-grams (n-gram range=3-5), Bangla pre-trained word embeddings, punctuation frequency, and normalized *Pars-of-Speech* frequency according to Hossain et al. (2020).

- Zhou (2017): employ a BiGRU (Cho et al., 2014) network with a self-attentive network (Yang et al., 2016) on top of the BiGRU representations and achieve the first position at Clickbait Challenge 2017 (Potthast et al., 2018a) with an F1 score of 0.683.

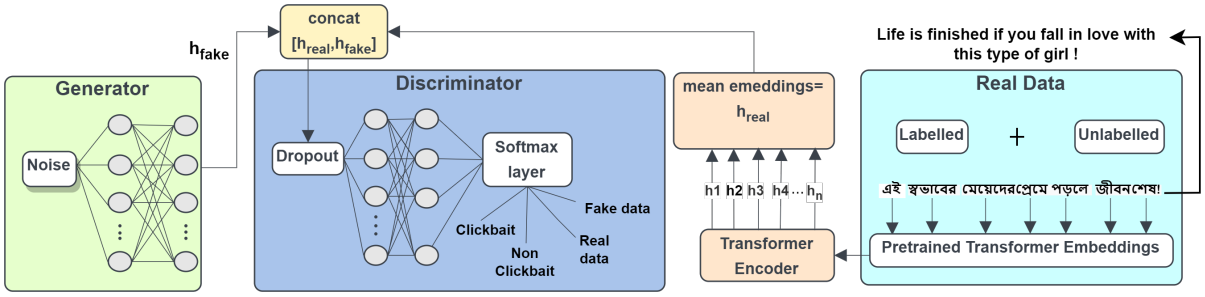


Figure 1: GAN-BanglaBERT architecture. Generator $\mathcal{G}$ generates fake data given random noise, and Discriminator $\mathcal{D}$ takes both real and this fake data and outputs four labels: 0 for non-clickbait, 1 for clickbait, 2 for real and 3 for fake data.

- Agrawal (2016): employ a multi-channel CNN model with one convolutional layer similar to the model demonstrated by Kim (2014). Pre-trained word embeddings are passed to multiple filters, and their concatenated representation is sent to a Max Pooling layer for the final representation.

- Lee et al. (2021): We translate all our article titles using a Bangla-to-English translator model Bangla-NMT (Hasan et al., 2020a) which outperformed Google Translate on SUPara-benchmark test set (Hasan et al., 2019). The translated titles are passed into a state-of-the-art misinformation detection model UnifiedM2 (Lee et al., 2021) trained on fake, clickbait, rumor, and news-bias datasets in English. We investigate if translating the titles and using a state-of-the-art model trained in English suffices for clickbait detection or whether language-specific training is necessary.

6 Experimental Setup

6.1 Pre Processing

Normalizer module by Hasan et al. (2020b) and Bangla unicode normalizer by (Alam et al., 2021) are used for Unicode and nukta normalization, removing HTML tags, URL links, etc. High punctuation usage is a common trait of clickbait titles. We preserve all syntactically correct punctuation in our titles and remove punctuation that appeared in the middle of words causing words to break and create out-of-vocabulary words for models.

6.2 Experimental Settings

For all models, we use the article's title as input as the title mainly creates the curiosity gap that is the principal characteristic of a clickbait title (Potthast et al., 2016). We use Bangla Fasttext (Bojanowski et al., 2017) and Bangla Word2Vec embedding pretrained on Bangla Wikipedia Dump Dataset with coverage of 65.16% and 60.91% respectively, on the total vocabulary size of article titles as embedding inputs. We extract the *Parts of Speech* (POS) tags using BNLP toolkit (Sarker, 2021). We derive a Bangla punctuation list from Alam et al. (2021). We experiment with both BiGRU and BiLSTM models for (Zhou, 2017) model and show the better performing one in section 7. The above models are trained for 40 epochs with Adam optimizer (Kingma and Ba, 2017) and learning rate = 2e-5, which is changed dynamically according to 1cycle learning rate scheduler (Smith and Topin, 2018). The GAN-BanglaBERT and BanglaBERT models are trained for 20 epochs with AdamW optimizer (Loshchilov and Hutter, 2019), and the learning rate is slowly increased from zero to 1e-5 within a warmup period. For GAN-BanglaBERT, the learning rate for the generator and discriminator model is kept the same. For all models, we pad or truncate titles to lengths of 64. The labeled dataset is split into 70:10:20 fashion for training, validation, and test splits using stratified sampling. All models are trained with batch size=64, and the best model based on the validation result is used to evaluate the final test set. Each experiment is repeated five times, and the average result on the held-out test set is used for the final result of all the models.

7 Results and Analysis

Table 4 illustrates the performance of all models on our test set. For each type of model, only the best-performing feature's result is shown. GAN-BanglaBERT outperforms all other models regarding F1 score, precision, metric, and recall. It achieves a 75.13% F1 score on the clickbait

Model	F1 Score	Precision	Recall	Accuracy
Zhou et al. (2016) (`Fasttext`)	39.37	39.88	38.87	57.87
Agrawal (2016) (`Fasttext`)	35.15	40.05	31.32	59.33
Logistic Regression (`character 3,4,5 gram`)	66.28	75.36	59.15	78.82
Random Forest (`character 3 gram`)	67.01	61.06	74.25	74.27
Lee et al. (2021)	11.02	39	6.4	63.53
BanglaBERT	71.72	80.42	64.71	82.04
GAN BanglaBERT	**75.13**	**75.45**	**74.81**	**82.57**
Human Baseline	76.81	77.6	76.04	77.01

Table 4: Performance comparison of GAN-BanglaBERT and all other models on the test set. F1 score, precision, and recall are for the clickbait class. For Zhou (2017) model, a better performing BiLSTM-attn model result is shown. GAN-BanglaBERT outperforms all other models and the performance difference is statistically significant ($p < 0.01$) according to McNemar's test (Dietterich, 1998)

class, which is 3.41% greater than the supervised BanglaBERT model. The performance is close to the human upper bound of 76.8% F1 score. The human baseline score shows that separating clickbait and non-clickbait titles is a difficult task even for humans, and clickbait may not be perceptible to all humans (Potthast et al., 2018b).

Figure 2 shows the ROC curve (receiver operating characteristic curve) for all models where the GAN-BanglaBERT model achieves the highest AUC (area under ROC curve) score of 0.8925, which is higher than the BanglaBERT. The high AUC score of GAN-BanglaBERT suggests that it can distinguish between clickbait and non-clickbait titles more accurately than other models.

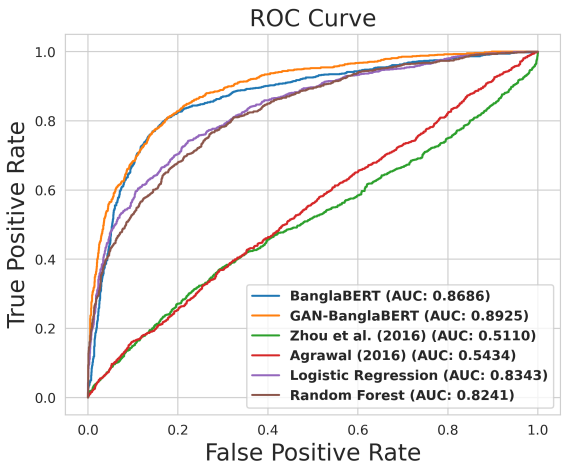


Figure 2: ROC curve for all models where GAN-BanglaBERT achieves the highest Area Under ROC Curve (AUC) score.

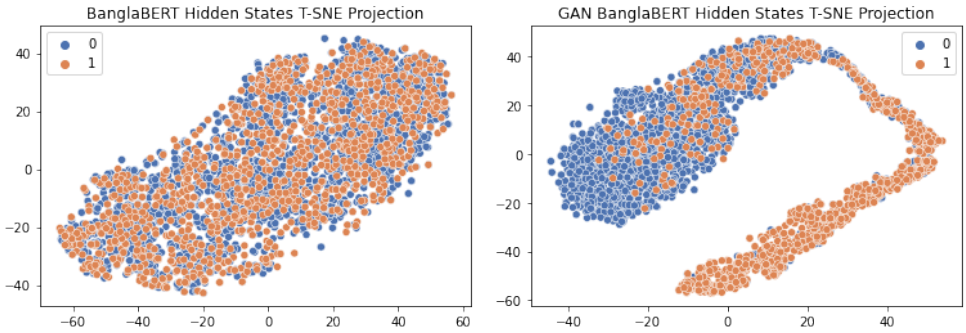
To investigate whether training in a semi-supervised approach improves BanglaBERT's inner representations as stated by (Croce et al., 2020), we plot the average of the last layer hidden representations of GAN-BanglaBERT and BanglaBERT using a t-SNE projection (van der Maaten and Hinton, 2008) in Figure 3b. GAN-BanglaBERT better separates the clickbait class from the non-clickbait than the BanglaBERT model, proving that training a BERT model in a semi-supervised adversarial manner can improve the learned representations of the model and thus improve performance.

For creating the unlabelled dataset, we choose clickbait-dense websites from the web to ensure a higher abundance of clickbait titles. To investigate whether this helps performance, we create another unlabelled dataset of the same size from Daily Prothom Alo archive[9], which has a substantially lower clickbait ratio. Our model achieves 72.38% F1 score on this second unlabelled set compared to 75.13% F1 score on the original unlabelled set, proving that a higher clickbait ratio in the unlabelled set improves performance on the Clickbait class.

Table 5 shows a prominent clickbait category - 'ambiguous' where GAN-BanglaBERT performs better than other models. 'They did not even forsake my mother! - Bhabna' is a quotation that implies something ostentatious happened with the mother, although expressed very vaguely. ছাড়ল না (not, forsake) words create this ambiguity which GAN-BanglaBERT correctly gives more attention to, but BanglaBERT fails to do so. The high AUC score and better separation in encoding shown in Figure 3 enables GAN-BanglaBERT to perform better in these harder-to-detect cases.

Table 4 shows that Lee et al. (2021) model on translated titles performs very poorly compared to

[9]https://github.com/zabir-nabil/bangla-news-rnn



(a) BanglaBERT t-SNE (b) GAN-BanglaBERT t-SNE

Figure 3: Visualization of last layer hidden representations using t-SNE for BanglaBERT (3a) and for GAN-BanglaBERT (3b). 0 represents Non-Clickbait and 1 represents Clickbait in both figures.

Category	Attention Weighted Words		Important Words
Ambiguous	BanglaBERT	এরা আমার মাকে ##ও ছাড়ল না : [UNK] ভাবনা	ছাড়ল, না (forsake, not)
	GAN-BanglaBERT	এরা আমার মাকে ##ও ছাড়ল না : [UNK] ভাবনা	
	Title	এরা আমার মাকেও ছাড়ল না: ভাবনা	
	Translation	They did not even forsake my mother	

Table 5: Comparison between GAN-BanglaBERT and BanglaBERT on ambiguous type clickbait title prediction. Each word is highlighted according to the attention weight given by the model.

other models. Machine translation produces more synthetic text, which diminishes the lexical and syntactical style and richness of the source language (Vanmassenhove et al., 2021). For example, অবিকল মানুষের মত করে দরদাম করে বাজারে ফল বিক্রি করছে বানর, তুমুল ভাইরাল ভিডিও

is translated to 'Monkeys selling fruit in the market at the expense of the real man, viral video.' Although this translation is factually correct, it loses the source language's exaggerated tone, leading to misclassification.

Logistic regression and Random Forest model on character TF-IDF features heavily outperform neural network models like BiLSTM with attention network and CNN (Zhou, 2017; Agrawal, 2016). These models can effectively identify certain keywords that are very significant in classifying clickbait titles. For instance, a top character feature returned by logistic regression is বললেন (told), which is a common keyword found in many clickbait titles, e.g., বড় সুখবর দিয়ে যা বললেন দীঘি (What Dighi said about the great news). The poor performance of neural network models can be attributed to Bangla pre-trained Fasttext and Word2Vec embeddings, which are trained on the Bangla Wikipedia dump and are significantly smaller in size than English. Training these embeddings on training data and then initializing the neural models with these embeddings may improve performance.

All models perform poorly on Bait & Switch type titles as mentioned in Table 1 where titles where the main content under-delivers the title's statements. As these types of clickbait require reading the content to predict correctly, all models underperform as they are trained on only the article's title. Effectively combining content features with titles to classify these types of clickbait titles is a future research endeavor for us.

8 Conclusion

We present the first clickbait detection dataset containing 15,056 labeled new articles and 65,406 unlabelled articles containing article title, content and metadata to enable researchers to use this dataset to build state-of-the-art clickbait detection models. By conducting a comprehensive study on various architectures, we provide a strong baseline for detecting clickbait in Bangla articles. We show that training a pre-trained Transformer model in a semi-supervised approach by incorporating unlabeled data improves performance and inner representation. As simple statistical models perform strongly on clickbait titles, we aim to investigate how these features can be combined with word embeddings to pass into neural networks. We also plan to investigate how features from article content can be utilized to detect clickbait. We wish to publicly release the dataset and code to further progress into Bangla clickbait detection.

Acknowledgments

We thank our data annotators and the volunteers from the Department of CSE, BRAC University, who participated in the human baseline experiment. We also thank the CSE Dept. of BRAC University, Bangladesh, for their continued support and direction. We show our sincerest gratitude to the anonymous reviewers and the pre-submission mentors for their valuable suggestions which helped improve the research work.

References

Amol Agrawal. 2016. Clickbait detection using deep learning. In *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, pages 268–272.

Samiul Alam, Tahsin Reasat, Asif Shahriyar Sushmit, Sadi Mohammad Siddique, Fuad Rahman, Mahady Hasan, and Ahmed Imtiaz Humayun. 2021. A large multi-target dataset of common bengali handwritten graphemes. In *International Conference on Document Analysis and Recognition*, pages 383–398. Springer.

Ron Artstein and Massimo Poesio. 2008. Survey article: Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596.

Abhik Bhattacharjee, Tahmid Hasan, Kazi Samin, Md Saiful Islam, M. Sohel Rahman, Anindya Iqbal, and Rifat Shahriyar. 2021. Banglabert: Combating embedding barrier in multilingual models for low-resource language understanding. *CoRR*, abs/2101.00204.

S Elizabeth Bird. 2008. Tabloidization. *The International Encyclopedia of Communication*.

Prakhar Biyani, Kostas Tsioutsiouliklis, and John Blackmer. 2016. " 8 amazing secrets for getting more clicks": detecting clickbaits in news streams using article informality. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Kyunghyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder-decoder approaches.

Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. Electra: Pre-training text encoders as discriminators rather than generators.

Danilo Croce, Giuseppe Castellucci, and Roberto Basili. 2020. GAN-BERT: Generative adversarial learning for robust text classification with a bunch of labeled examples. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2114–2119, Online. Association for Computational Linguistics.

Thomas G. Dietterich. 1998. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation*, 10(7):1895–1923.

Manqing Dong, Lina Yao, Xianzhi Wang, Boualem Benatallah, and Chaoran Huang. 2019. Similarity-aware deep attentive model for clickbait detection. In *Advances in Knowledge Discovery and Data Mining*, pages 56–69, Cham. Springer International Publishing.

Jeffrey L Elman. 1990. Finding structure in time. *Cognitive science*, 14(2):179–211.

J.L. Fleiss et al. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.

Abhijnan Chakraborty; Bhargavi Paranjape; Sourya Kakarla; Niloy Ganguly. 2016. Stop clickbait: Detecting and preventing clickbaits in online news media. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial networks.

Md. Arid Hasan, Firoj Alam, Shammur Absar Chowdhury, and Naira Khan. 2019. Neural machine translation for the bangla-english language pair. In *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, pages 1–6.

Tahmid Hasan, Abhik Bhattacharjee, Kazi Samin, Masum Hasan, Madhusudan Basak, M. Sohel Rahman, and Rifat Shahriyar. 2020a. Not low-resource anymore: Aligner ensembling, batch filtering, and new datasets for bengali-english machine translation.

Tahmid Hasan, Abhik Bhattacharjee, Kazi Samin, Masum Hasan, Madhusudan Basak, M. Sohel Rahman, and Rifat Shahriyar. 2020b. Not low-resource anymore: Aligner ensembling, batch filtering, and new datasets for Bengali-English machine translation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2612–2623, Online. Association for Computational Linguistics.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Md Zobaer Hossain, Md Ashraful Rahman, Md Saiful Islam, and Sudipta Kar. 2020. Banfakenews: A dataset for detecting fake news in bangla. *arXiv preprint arXiv:2004.08789*.

Vijayasaradhi Indurthi, Bakhtiyar Syed, Manish Gupta, and Vasudeva Varma. 2020. Predicting clickbait strength in online social media. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4835–4846.

Yoon Kim. 2014. Convolutional neural networks for sentence classification.

Diederik P. Kingma and Jimmy Ba. 2017. Adam: A method for stochastic optimization.

J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Nayeon Lee, Belinda Z. Li, Sinong Wang, Pascale Fung, Hao Ma, Wen-tau Yih, and Madian Khabsa. 2021. On unifying misinformation detection.

Vladimir I. Levenshtein. 1965. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics. Doklady*, 10:707–710.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization.

Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. *arXiv preprint arXiv:1310.4546*.

Mahmud Hasan Munna and Md Shakhawat Hossen. 2021. Identification of clickbait in video sharing platforms. In *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)*, pages 1–6.

Martin Potthast, Tim Gollub, Matthias Hagen, and Benno Stein. 2018a. The clickbait challenge 2017: Towards a regression model for clickbait strength. *arXiv preprint arXiv:1812.10847*.

Martin Potthast, Tim Gollub, Kristof Komlossy, Sebastian Schuster, Matti Wiegmann, Erika Patricia Garces Fernandez, Matthias Hagen, and Benno Stein. 2018b. Crowdsourcing a large corpus of clickbait on twitter. In *Proceedings of the 27th international conference on computational linguistics*, pages 1498–1507.

Martin Potthast, Sebastian Köpsel, Benno Stein, and Matthias Hagen. 2016. Clickbait detection. In *European Conference on Information Retrieval*, pages 810–817. Springer.

Md Main Uddin Rony, Naeemul Hassan, and Mohammad Yousuf. 2017. Diving deep into clickbaits: Who use them to what extents in which topics with what effects? In *Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017*, pages 232–239.

Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. 2016. Improved techniques for training gans.

Sagor Sarker. 2021. Bnlp: Natural language processing toolkit for bengali language.

Leslie N. Smith and Nicholay Topin. 2018. Super-convergence: Very fast training of neural networks using large learning rates.

Philippe Thomas. 2017. Clickbait identification using neural networks.

Eva Vanmassenhove, Dimitar Shterionov, and Matthew Gwilliam. 2021. Machine translationese: Effects of algorithmic bias on linguistic complexity in machine translation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2203–2213, Online. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *ArXiv*.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1480–1489, San Diego, California. Association for Computational Linguistics.

Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. 2016. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 207–212, Berlin, Germany. Association for Computational Linguistics.

Yiwei Zhou. 2017. Clickbait detection in tweets using self-attentive network.

A Appendix

A.1 Data sources

We choose a site for scraping if the homepage seems to contain more clickbait than non-clickbait titles after a cursory glance by the annotators. We also select some famous Bengali online news publishers such as Kaler Kantha4, SomoyTV5, and

RTV news6 for scraping to facilitate future investigation into clickbait practices in popular Bangla news mediums. Table 6 contains the data sources for our dataset's labeled and unlabelled portions.

Labelled		
Domain	News Count	
	clickbait	non-clickbait
twentyfourbd	1727	1062
topdhaka	1004	920
rtvonline	1003	86
BanFakeNews	623	415
kureghornews	634	375
newzcitizen	633	351
nbtimes24	750	228
citynewszet	503	451
authoritynewz	561	385
thebasenewz	537	268
newzauthority	308	281
newsholder21	361	117
techzoom	369	60
channeldhaka	291	38
kalerkantho	285	37
somoynews	194	38
beanibazarview24	109	52

	Domain	News Count
Unlabelled	mtnews24	16194
	dakpeon24	1567
	newsfastcreator	8836
	propernewsbd	1830
	thecityvpn	14455
	usbanglanews	16099
	glamourbd	6197
	jagonews24	228

Table 6: Data sources of Bangla Clickbait Datset

A.2 Detailed Results

For statistical models, we experimented with Random Forest and Logistic Regression networks. We passed various types of lexical, syntactical, and embedding features to these networks to investigate which performs best. For neural network models, we employ architectures from two previous research works Zhou et al., 2016; Agrawal, 2016. For Transformer networks, we train commonly available Bangla pre-trained transformer models in both classic and semi-supervised GAN manner. Table 7 contains the results of these experiments.

Statistical Classifiers		
Traditional Linguistic Features	**Logistic Regression**	**Random Forest**
Unigram (U)	57.39	56.11
Bigram (B)	29.7	53.34
U+B+T	57.68	55.94
C-3 gram	64.81	**67.01**
C-4 gram	65.59	62.48
C-5 gram	65.13	58.58
C3+C4+C5	**66.28**	65.36
All Lexical(L = U+B+T+C3-C5)	64.29	65.6
Parts of Speech(POS)	33.14	40.37
L+POS	62.23	65.97
Embedding Word2Vec (E W)	53.11	51.35
Embedding Fasttext (E F)	50.19	49.4
L+POS+E W	64.2	65.04
L+POS+E F	64.43	65.05
Punctuation (P)	5.88	52.06
L+POS+E W+P	63.34	64.7
L+POS+E N+P	64.34	64.91
All features	64.73	63.26

Transformer Networks	**Classic**	**SS-GAN**
BERT base multilingual cased	62.37	70.21
Bangla BERT Base	68.13	68.54
Indic-BN-BERT	72.21	73.36
Indic-BN-RoBERTa	67.76	70.52
DistilBERT base multilingual cased	69.61	70.38
Indic-BN-DistilBERT	71.32	72.35
Bangla-Electra	66.79	67.77
Indic-BN-XLM-RoBERTa	71.82	70.75
CSENLP-BanglaBert	**71.72**	**75.13**
CSENLP-BanglaBert_Large	71.66	72.07
Neural Networks		
CNN (Agrawal, 2016)	35.15	
Bi-LSTM (Zhou et al., 2016)	39.37	

Table 7: Detailed result of all experiments conducted on BanglaBait dataset

A.3 Difference between Clickbait and Fake news

Although Hossain et al. (2020) published the first dataset of Bangla Fake news, we don't focus on the misinformation, fabricated or fake content within the articles, or their authenticity to detect clickbait in this dataset. The following two examples explain the difference between fake and clickbait titles in detail-

Example 1: Buying land on the Moon is the current craze. Explore how you can do that too!
চাঁদে জমি কিনার হিড়িক, জেনে নিন আপনিও কিভাবে কিনবেন

Example 2: 'Hawa' got nominated for the Oscars
অস্কারে মনোনয়ন পেয়েছে 'হাওয়া'

Example 1 presents an accurate title (verified by renowned news publishers such as the Kalerkantho and the Somoynews) in a clickbait-style by using hyperbolic words like 'craze' and alluring phrases like 'Explore how you can do that too.' It proves a clickbait article does not have to be fake to be clickbait. Example 2, on the other hand, is fake news verified from the official Facebook page of the movie 'Hawa', however, the title style is not exactly luring readers to click, proving that an article can be fake without being clickbait. In short, clickbait headlines do not necessarily have to be fake news; they may contain genuine information but in an exaggerated fashion (Dong et al., 2019). Biyani et al. (2016) includes factually wrong articles in the 'wrong' category of clickbait articles.