CLASSIFYING BANGLA HEALTH MISINFORMATION FROM SOCIAL MEDIA USING MACHINE LEARNING

PARINDA RAHMAN[1], EMON SARKER[1], MAHIMA AHSAN[1], INTISAR TAHMID NAHEEN[1], FERDOUS BIN ALI[2]

[1]Department of Electrical and Computer Engineering, North South University, Dhaka, Bangladesh
[2]Statistics Department, Jahangirnagar University, Dhaka, Bangladesh
E-MAIL: [1]parinda.rahman@northsouth.edu, [1]emon.sarker@northsouth.edu, [1]mahima@ahsan@northsouth.edu,
[1]intisar.naheen@northsouth.edu, [2]hridoyferdous@yahoo.com

Abstract:

The growing use of social media platforms offers a means to spread health-related information at an increased rate. However, disseminating health information on social media may be hazardous as the information often has no regulation. Therefore, it is crucial to find ways to classify health misinformation. Using machine learning models, this project utilizes mined Bangla text data and its English Translations to classify health misinformation from mined social media data. Multiple models were used such as Random Forest, XGBoost, SVC, etc, The highest reported precision in Bangla text and the English translation is 77% and 79% respectively. XGB classifier had the highest accuracy of 75% for English translation and the Extra Trees Classifier had the highest accuracy of 72% for Bangla text.

Keywords:

Health, Medicine, Social Media, Bangladesh, Machine Learning.

1. Introduction

Over time social media became a popular outlet to share health information [1]. Health organizations use this platform to disseminate health information [2] however these public platforms have unprecedented social and health risks [3]. Studies suggest a higher likelihood of false and misleading information spreading through social media than scientific knowledge [4]. Health misinformation is defined as an assertion that cannot be scientifically supported and can be deceptive [5]. The Kepois analysis' indicated a 10.1 percent increase in social media users in Bangladesh between 2021 and 2022 [6]. Furthermore, it has been observed that the consumption of social media increased rapidly during the Covid-19 pandemic [7]. Before the COVID-19 epidemic, misleading promises of miracle treatments and conspiracy theories about the ”hidden” causes of many chronic health disorders had a harmful impact on sufferers [8]. Parallel to the COVID-19 pandemic, a ”huge infodemic” has been declared by the World Health Organization due to the significant rise in false information [9]. Moreover, health misinformation on social media caused between 2 to 12 million to be unvaccinated because of vaccine hesitancy [10]. Therefore, it is crucial to ensure effective ways of detecting health misinformation.

One main area of focus in research has been the application of machine learning methods to combat the propagation of false information. A twenty-three supervised machine learning model was used by [11] to detect misinformation in a study. A huge quantity of data on social media cannot be moderated through human intervention, therefore well-performing models are crucial to combat misinformation. While multiple studies were done to provide techniques to combat health misinformation during COVID-19, they focused on English text [12], and measures were taken to remove those posts; however, a large majority of Bengali text remained unmoderated. A study developed a model for the detection of fake news using a Bangla natural language processing model, however, it was not implemented for health misinformation or using machine learning techniques [13]. As a result, misleading health information is rampant on social media using Bangla text.

Therefore, the paper implements a classification technique for health misinformation using Bengali text and its English translations using both machine learning. Our contributions to this work include:

- A novel dataset of text data was curated using manual data scraping from multiple platforms such as Facebook, YouTube, etc
- Another significant contribution of this work is that it applies multiple machine learning models to classify health misinformation in Bangla and English translation and then compares its performance.
- A web application has been developed using our best-performing model for people to fact-check valid and misleading information using both Bangla and English text.

2. Literature Review

2.1. Health Misinformation

Health misinformation is a crucial challenge that needs to be combated on different social media platforms because studies suggest that health misinformation spreads more easily than proven scientific knowledge. A systematic review that used empirical findings to assess the magnitude of health misinformation spread focused on six major health domains such as vaccines, drugs or smoking, pandemics, eating disorders, and so on. Their findings showed that smoking products and drugs had the highest rate of misinformation across different studies [14]. In literature [15], the word ”misinformation” is used interchangeably with similar-meaning words like ”disinformation”, ”fake news”, etc. The findings of the literature revealed that the general public's anxiety and uncertainty increase because of misinformation.

2.2. Machine Learning as a solution

Machine learning tools allow organizations and agencies to monitor misinformation on large social media platforms. A study uses a machine learning model to conduct four separate studies [16]. The model could keep track of multiple cases of misinformation at the same time, with an F1 score greater than 87%. The models used were K-Nearest Neighbors, Decision Trees, Random Forests, XGBoost, AdaBoost, Support Vector Machines, and Multilayer Perceptrons. Support vector machines had the highest macro-average F1 score 87.2% score in their study. As there is an enormous amount of misinformation being propagated online, it is difficult to counter the volume of misinformation. Therefore, another study [17] used a machine-learning (ML) based misinformation detection model about COVID-19. The dataset only used Twitter data on the public CoAID misinformation dataset.). The best performance was observed on KNN (where k = 3). Another study [18] used both ML and Deep Learning for classifying fake news and found an average accuracy of 94.21% using LSTM. Good performance has been observed for most public datasets. A study [19] achieved 99.8% and 100% accuracy on the Liar and ISOT datasets respectively. The accuracy is better in comparison to the previously stated literature [19] because they use cutting-edge machine learning models. Nevertheless, it is important to use these models on raw datasets. Therefore, previous literature suggests that Machine learning is a useful technique for combating misinformation.

2.3. Mined Datasets & Machine Learning

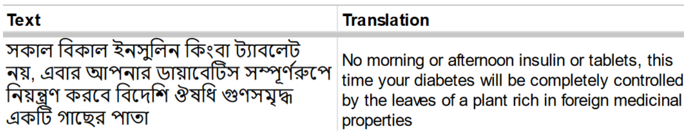
Some studies use raw datasets for the classification and detection of fake information on social media. Using credible sources and validation from medical professionals, 15,000 tweets on vaccines were stated in the paper [20]. The Parsehub tool was used to scrape the content of datasets. An intelligent augmentation method is employed to increase the amount of bogus news in the dataset. The proposed method used linear regression to reach 91% accuracy.

WhatsApp conversations that were manually categorized and collected from public groups in Brazil were used in a study [21]. Moreover, a total of 108 classifiers for identifying misinformation were employed, utilizing a blend of natural language processing-based feature extraction techniques and commonly used machine learning algorithms. The F1 score was 0.87 when texts with fewer than 50 words were filtered. Another study used a dataset collected from different Facebook profiles [22]. It used data from fact-checking software to detect fake news. Therefore, classification tasks can be performed using raw datasets.

3. Methodology

3.1 Text data collection and classification

Text data from public posts were collected from the social media platforms such as Facebook and YouTube by utilizing each of their respective search tools. The queries entered into the search tool were done manually as the API of these social media platforms does not allow gathering content made by user accounts. The 1500 text data collected was split into two classes: Valid Information and Misinformation. There are 796 valid information data points and 699 misinformation data points. An example of the data can be seen in Fig. 1. The translation for the data was generated using the Google Translate API and then the data was manually verified by native Bengali language users to ensure the meaning and context of the text remained intact.



Text	Translation
সকাল বিকাল ইনসুলিন কিংবা ট্যাবলেট নয়, এবার আপনার ডায়াবেটিস সম্পূর্ণরূপে নিয়ন্ত্রণ করবে বিদেশি ঔষধি গুণসমৃদ্ধ একটি গাছের পাতা	No morning or afternoon insulin or tablets, this time your diabetes will be completely controlled by the leaves of a plant rich in foreign medicinal properties

FIGURE 1. Example of original Bangla text that has been extracted and its translation

3.2 Data Preprocessing

The text data was then preprocessed to prepare it for modeling. First, any URLs, links, and punctuations were removed by using a regular expression (regex) script. Afterward, the stopwords were removed using the Natural Language Toolkit (NLTK). Removing stopwords is an important and necessary step in any machine-learning task regarding Natural Languages. Stopwords are the most common words in any language that do not add much information to the text they appear in. These could be pronouns, conjunctions, or prepositions. Removing them will reduce the dimensionality of our text data. In order to work with both the Bangla text and the English translation of it, the stopword removal had to be performed separately for both the Bangla text and the English translation. The English stopwords were removed with the help of the NLTK library. The Bangla stopwords were removed by modifying the NLTK library with an external array of Bangla stopwords.

The text data was converted into a numerical representation using the CountVectorizer from the Sklearn (Scikit-learn) library. The CountVectorizer tokenizes each word and counts the frequency of each unique token found in the text data. In addition to converting text data into a numerical representation, CountVectorizer emphasizes word frequency and domain-specific words which will prove helpful if the discriminatory feature between valid information and misinformation is the vocabulary used. The max_features parameter of the CountVectorizer was set to 250 which resulted in the best performance during testing. The label was also encoded.

3.3 Machine Learning Models

The encoded data was split into training and test sets. Afterward, various machine learning models were trained to classify the data into two categories: valid information and misinformation. Classification and Regression Trees (CART) based models such as Extra Trees Classifiers and Extreme Gradient Boosting (XGB) algorithm are particularly good at classifying text data as it is capable of modeling high-dimensional and sparse data [23].

The Extra Trees Classifier is a modification of the Random Forest Classifier. In the decision trees in Extra Trees Classifier, the thresholds of discrimination are randomized. The best features are selected by first ordering the random features in ascending order based on the Gini calculation of each feature [24].

$$Gini = 1 - \sum_{i}^{n} (\mu_i)^2 \tag{1}$$

The XGB classifier utilizes gradient boosting to combine the ensemble of CARTs which allows it to achieve better performance. The XGB classifier loss function used is given below:

$$L = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(X_i)) + \Omega(f_t) \tag{2}$$

Additionally, multiple machine learning algorithms were also trained. These classifier models include extreme Gradient Boosting (XGB), Light Gradient Boosting Machine (LGBM), Random Forest, Extra Trees, Support Vector Machine (SVC), AdaBoost, Ridge, Bernoulli Naive Bayes, K-Nearest Neighbors, Bagging classifier, and more. The models were assessed based on various metrics, out of which we closely observe the precision scores as a medical information classification task should emphasize the reduction of false positives.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{3}$$

3.4 Project Workflow

Fig. 2 shows the steps employed in the pipeline. The data was initially collected and preprocessed and passed to various models and the results were evaluated.

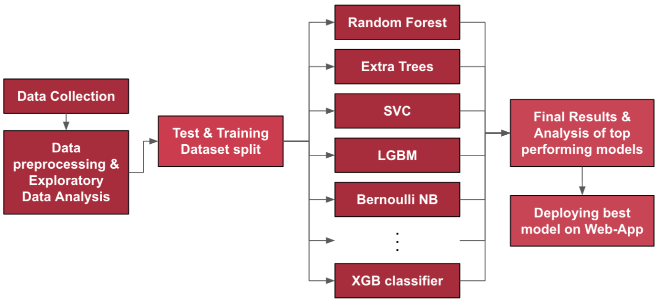
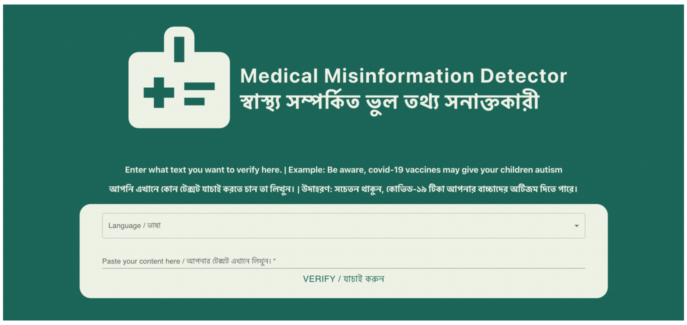

Data Collection
Data preprocessing & Exploratory Data Analysis
Test & Training Dataset split
Random Forest
Extra Trees
SVC
LGBM
Bernoulli NB
⋮
XGB classifier
Final Results & Analysis of top performing models
Deploying best model on Web-App

FIGURE 2. Workflow diagram of the entire process

3.5 Web Application

The best-performing model was pickled and used for deployment in a web application. This is a web application that utilizes machine learning to classify a text input as either misinformation or valid information. The user interface of the web application is shown in Fig. 2. The pickled model was deployed as a simple API built using Flask, a Python-based framework to build APIs. The front end is a single-page application created using React.js. It gives the user two language options through a drop-down menu of "English" or "Bangla". Through passing the input data through our model a prediction is given to the user through a pop-up. As a precaution, the pop-up encourages the user to consult a medical professional to be completely sure of whether or not a piece of medical information is accurate. For the time being, the web app is hosted on the local server.


Medical Misinformation Detector
স্বাস্থ্য সম্পর্কিত ভুল তথ্য সনাক্তকারী

Enter what text you want to verify here. | Example: Be aware, covid-19 vaccines may give your children autism

আপনি এখানে কোন টেক্সট যাচাই করতে চান তা লিখুন। | উদাহরণ: সচেতন থাকুন, কোভিড-১৯ টিকা আপনার বাচ্চাদের অটিজম দিতে পারে।

FIGURE 3. Screenshot of the web application

4. Results and Discussion

We evaluate our results based on metrics such as accuracy, precision, recall, and f1-scores. We trained multiple Machine Learning models and reported the Top-5 well-performing models in Table I. The comparison of precision in Bangla and English text is shown in Fig 4. Due to the nature of our classification problem which deals with medical data, we focus more on the precision scores that evaluate true positives. The extra trees classifier showed the highest precision of 79% and 77% on both English translations and Bangla texts respectively. In terms of accuracy, The XGBoost Classifier showed the highest performance with 75% accuracy for English Translations and the Extra Trees Classifier showed the highest performance with 72% accuracy for Bangla texts. These reported accuracy scores may have been a consequence of having a small dataset that only consists of around 1500 data points. The overall accuracy of Bangla texts being lower than English Translations could be due to the lexical complexity of the language.

TABLE 1. Top-5 models for English Translations and Bangla Texts

Text	Model	Accuracy	Precision	Recall	F1-score
English Translations	XGB Classifier	75%	76%	72%	75%
	Random Forest Classifier	74%	78%	66%	74%
	SVC	73%	77%	67%	73%
	Extra Trees Classifier	73%	79%	62%	73%
	Logistic Regression	73%	74%	70%	73%
Bangla	Extra Trees Classifier	72%	77%	62%	72%
	Random Forest Classifier	72%	75%	62%	72%
	XGB Classifier	71%	73%	68%	71%
	LGBM Classifier	71%	73%	66%	71%
	SVC	70%	73%	63%	70%

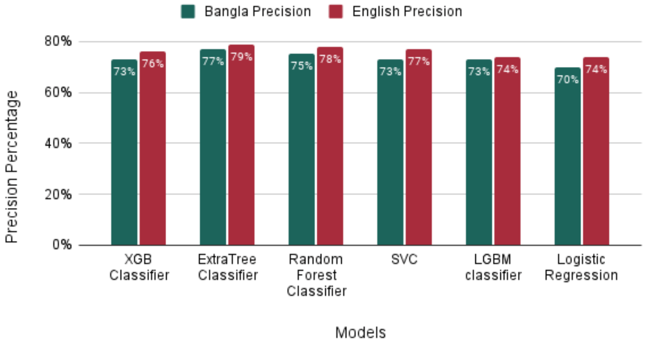


FIGURE 4. Precision comparison between models with Bangla and English texts

Due to the size of our dataset, models that utilize bootstrapping such as the Random Forest classifier, extra trees classifier, or the XGB classifier outperformed the Support Vector Classifier (SVC). The extra trees classifier gave us the highest precision scores for both English translations and Bangla texts. This can be due to the nature of the algorithm which escapes bias by sampling the entire dataset, instead of sampling subsets using bootstrapping as done in random forest classifiers. Most of these models also utilize ensemble learning within the algorithm which helps escape the issue of overfitting.

It is interesting to note that tree-based classifiers such as Extra Trees, and XGB classifiers turned out to be the best-performing models, it is suggestive that the mined data could have high levels of noise. These classifiers are quite robust when modeling noisy data [25]. Another contributing factor to their high performance could be the use of a CountVectorizer as it transforms the raw data into a bag-of-words, which is a high-dimensional, sparse vector representation. These tree-based classifier models can handle such data quite well [26].

As tree-based classifiers have performed well, there is a good chance a deep learning-based classifier will perform even better as it will be able to classify based on the semantic meaning of the text. Semantic-based classification may give better results in such cases [27]. This, in the long run, will prove beneficial as the difference between valid information and misinformation is not in the lexical or vocabulary differences between texts, but rather the semantics it portrays.

5. Conclusions and Future Work

The capability of the models suggests that machine learning-based approaches are effective in classifying health misinformation from social media data. Furthermore, when compared to other recent literature, our model's precision is comparatively lower which is an issue as it is dealing with the classification of medical misinformation.

In the future, the raw dataset can be expanded with more data points and the model could be retrained using deep learning techniques. However, for a pilot attempt, the fact that a small raw dataset of fewer than 2000 data points was able to achieve a decent classification performance is promising. It is indicative that classifying Bangla medical misinformation can be achieved with a larger training dataset. Automated scraping tools may be used in the future for faster data mining and collection to improve the dataset. Furthermore, instead of a web app, a web extension of the model could be used for mass usage. A web extension will be able to classify, highlight and warn users about medical misinformation actively as they browse the internet. This would be a more effective tool to combat medical misinformation online.

Acknowledgements

We would like to extend our gratitude towards Dr.Ashrafee Afreen for her support in the verification of the dataset.

References

[1] W.-Y. S. Chou, A. Oh, and W. M. Klein, "Addressing health-related misinformation on social media," JAMA, vol. 320, no. 23, p. 2417, 2018.

[2] F. Xiong and Y. Liu, "Opinion formation on social media: An empirical approach," Chaos: An Interdisciplinary Journal of Nonlinear Science, vol. 24, no. 1, p. 013130, 2014.

[3] D. N. Cavallo, W.-Y. S. Chou, A. McQueen, A. Ramirez, and W. T. Riley, "Cancer prevention and control interventions using social media: User-generated approaches," Cancer Epidemiology, Biomarkers & Prevention, vol. 23, no. 9, pp. 1953–1956, 2014.

[4] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146–1151, 2018.

[5] S. Funk, M. Salathé, and V. A. Jansen, "Modelling the influence of human behaviour on the spread of infectious diseases: A Review," Journal of The Royal Society Interface, vol. 7, no. 50, pp. 1247–1256, 2010.

[6] DataReportal (2022), "Digital 2022 Bangladesh,Available:https://datareportal.com/reports/digital-2022-bangladesh

[7] "India: Covid-19 impact on media consumption by type of media 2020," https://coderwall.com/p/wntyia/how-to-cite-a-website-in-latex, 2021.

[8] Grimes, D. R. (2021). Medical disinformation and the unviable nature of covid-19 conspiracy theories. PLOS ONE, 16(3). https://doi.org/10.1371/journal.pone.0245900

[9] World Health Organization, rep., Feb. 2022. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf.

[10] R. Bruns, T. K. Sell, M. Trotochaud, and D. Hosangadi, "Covid-19 vaccine misinformation and disinformation costs," Johns Hopkins

Center for Health Security. [Online]. Available: https://www.centerforhealthsecurity.org/our-work/publications/covid-19-vaccine-misinformation-and-disinformation-costs-an-estimated-50-to-300-million-each-da. [Accessed: 06-Dec-2022].

[11] F. A. Ozbay and B. Alatas, “Fake news detection within online social media using supervised artificial intelligence algorithms,” Physica A: Statistical Mechanics and its Applications, vol. 540, p. 123174, 2020.

[12] Z. Wang, Z. Yin and Y. A. Argyris, ”Detecting Medical Misinformation on Social Media Using Multimodal Deep Learning,” in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 6, pp. 2193-2203, June 2021, doi: 10.1109/JBHI.2020.3037027

[13] A. Aggarwal, A. Chauhan, D. Kumar, M. Mittal, and S. Verma, “Classification of fake news by fine-tuning deep bidirectional transformers based language model,” ICST Transactions on Scalable Information Systems, p. 163973, 2018.

[14] V. Suarez-Lledo and J. Alvarez-Galvez, “Prevalence of health misinformation on social media: Systematic review,” Journal of Medical Internet Research, vol. 23, no. 1, 2021.

[15] YJ. Li, J. J. Marga, C. MK Cheung, XL. Shen, M. Lee, “Health Misinformation on Social Media: A Systematic Literature Review and F view and Future Research Directions” AIS Transactions on Human-Computer Interaction, vol. 14, no. 2, 2022

[16] K. Hunt, P. Agarwal, and J. Zhuang, “Monitoring misinformation on Twitter during crisis events: A machine learning approach,” Risk Analysis, vol. 42, no. 8, pp. 1728–1748, 2020

[17] M. N. Alenezi and Z. M. Alqenaei, “Machine learning in detecting COVID-19 misinformation on Twitter,” Future Internet, vol. 13, no. 10, p. 244, 2021.

[18] S. A. Alameri and M. Mohd, ”Comparison of Fake News Detection using Machine Learning and Deep Learning Techniques,” 2021 3rd International Cyber Resilience Conference (CRC), 2021, pp. 1-6, doi: 10.1109/CRC50527.2021.9392458.

[19] S. Hakak, M. Alazab, S. Khan, T. R. Gadekallu, P. K. Maddikunta, and W. Z. Khan, “An ensemble machine learning approach through effective feature extraction to classify fake news,” Future Generation Computer Systems, vol. 117, pp. 47–58, 2021.

[20] K. Hayawi, S. Shahriar, M. A. Serhani, I. Taleb, and S. S. Mathew, “ANTi-vax: A novel twitter dataset for COVID-19 vaccine misinformation detection,” Public Health, vol. 203, pp. 23–30, 2022

[21] L. Cabral, J. Monteiro, J. Franco da Silva, C. Mattos, and P. Mourão, “FakeWhastApp.BR: NLP and machine learning techniques for misinformation detection in Brazilian portuguese whatsapp messages,” Proceedings of the 23rd International Conference on Enterprise Information Systems, 2021.

[22] S. R. Sahoo and B. B. Gupta, “Multiple features based approach for automatic fake news detection on social networks using Deep Learning,” Applied Soft Computing, vol. 100, p. 106983, 2021.

[23] Dhieb, N., Ghazzai, H., Besbes, H., & Massoud, Y. (2019). Extreme gradient boosting machine learning algorithm for Safe Auto Insurance Operations. 2019 IEEE International Conference on Vehicular Electronics and Safety (ICVES). https://doi.org/10.1109/icves.2019.8906396

[24] Thankachan, K. (2022, August 9). What? when? how?: Extratrees Classifier. Medium. Retrieved March 18, 2023, from https://towardsdatascience.com/what-when-how-extratrees-classifier-c939f905851c

[25] Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., &amp; Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing, 67, 93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

[26] Reddy, G. T., Reddy, M. P., Lakshmanna, K., Kaluri, R., Rajput, D. S., Srivastava, G., &amp; Baker, T. (2020). Analysis of dimensionality reduction techniques on Big Data. IEEE Access, 8, 54776–54788. https://doi.org/10.1109/access.2020.2980942

[27] Saidani, N., Adi, K., &amp; Allili, M. S. (2020). A semantic-based classification approach for an enhanced spam detection. Computers &amp; Security, 94, 101716. https://doi.org/10.1016/j.cose.2020.101716