Novelty Journals

ISSN 2394-7314

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 10, Issue 3, pp: (7-15), Month: September - December 2023, Available at: www.noveltyjournals.com

Empirical Perspective on Significant Technological Barriers in Detecting Multilingual Cyberbullying in Nigerian Languages

[1]Christiana A. Okoloegbo, [2]Udoka F. Eze, [3]Gloria A. Chukwudebe, [4]Obi C. Nwokonkwo

[1]Department of Cybersecurity, Federal University of Technology, Owerri, Nigeria

[2]Department of Information Technology, Federal University of Technology, Owerri, Nigeria

[3]Department of Electrical and Electronics Engineering, Federal University of Technology, Owerri, Nigeria

[4]Department of Information Technology, Federal University of Technology, Owerri, Nigeria

DOI: https://doi.org/10.5281/zenodo.10008531

Published Date: 16-October-2023

Abstract: **In recent years, the potential for cyberbullying has grown in lockstep with the growth of social media. Cyberbullying has increased in Nigeria, where 49.6% (104.4 million) of Nigerians have online access and 26.14 percent of African internet users, predominantly Nigerians, have internet access. Cyberbullying occurs in many languages, not just English. Nigeria, on the other hand, has over 525 native languages, making it one of the world's most linguistically varied countries. Cyberbullying thrives in Nigeria due to the large audience, hence it is vital to identify cyberbullying in multiple languages. However, few researchers addressed this puzzle, hence this paper studied the technological barriers to multilingual cyberbullying detection in Nigerian local languages. An empirical perspective of descriptive survey was used to elicit expert-based responses from 60 purposively sampled cybersecurity engineers with 5 years and above expertise in machine learning and cyberbullying detection across five geopolitical zones in Nigeria to actualize the study's purpose. For variable dimension reduction, Principal Component Analysis (PCA) was used, and linear multiple regression was used to predict the model. The study's findings clearly show that the primary technological variables impeding cyberbullying detection in both English and any of Nigeria's primary indigenous languages of Hausa, Yoruba, and Igbo are language ambiguity, sarcasm dialect classification, domain influence, data scarcity, non-trusted source of data, difficulty in developing novel classifiers for multilingual text detection, annotation of training data, and language characteristics. It is recommended that study be performed to discover latent cyberbullying actions that cannot be identified in English Languages.**

Keywords: **Barriers, Cyberbullying, Detection, Multilingual, Nigerian Languages, Technology.**

I. INTRODUCTION

For young people, the internet has become a crucial development tool. It is an excellent source of knowledge as well as a communication tool. Despite its many advantages, the Internet may serve as a breeding ground for bullying [1]. Moreover, the use of information and communication technology for the abuse or mistreatment of another is referred to as

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 10, Issue 3, pp: (7-15), Month: September - December 2023, Available at: www.noveltyjournals.com

cyberbullying [2]. It is often perpetrated against a distant victim, yet in certain cases the perpetrator(s) may be close to the victim [3]. Cyberbullying takes many forms on social networking sites, including threats, harassment, and exploiting prospective victims [4].

Nigeria is a rich, diversified, and promising country, yet it has unique obstacles, as does any other country. One of the issues that young people face is cyberbullying. It has the potential to progressively expand over time, owing to rising internet penetration and the availability of more affordable smartphones [5]. Unfortunately, because these of the techniques employed, the expansion of social media and messaging platforms has also contributed [6]. Hence, cyberbullying is a rising problem among Internet users in Nigeria, and most people are unaware of it [7].

With the exception of the victims, cyberbullying identification is viewed as a tough and complex procedure, even for humans, due to the variations in the language used [8]. As a result, it is a major challenge that is attracting academics from all around the world to try to figure out if a post or a tweet is offensive or not. Most used methods for automatically detecting cyberbullying rely on English text and related forums [9]. However, in Nigeria, the bulk of cyberbullying is hate speech motivated by tribal insults [10]. According to Nairaland [11], most Yoruba locals on social media find it difficult to correct someone without using the pejorative suffix. For example, "do it like this ode (fool), click the red button, oponu (idiot), hold it for me, didirin (fool), you can't greet somebody, alaileko (lack of home instruction)," and so on. Nigerians of all tribes are notorious for using disparaging words, and thus no research work has been developed to create a system that automatically detects cyberbullying both in English text and Nigerian language text. This sheer research negligence necessitates the relevance to empirically identify and evaluate all the critical technological factors affecting cyberbullying detection in other languages in Nigeria.

II. RELATED WORKS

Ghosh et al. [12] did a Bengali language study on Social Media Cyberbullying Detection Using Machine Learning. Machine learning may be used to identify the linguistic designs used by threats and establish algorithms to recognize digitally harassing content. Most suggested studies on cyberbullying detection using machine learning have been written in English, Chinese, and Arabic. Regional Indian languages have received very little attention. They presented a methodology in their study that detects cyberbullying content in a rare or regional Indian language like Bengali. Furthermore, Talpur et.al [13] conducted a study on cyberbullying detection in Roman Urdu Language using Lexicon Based Approach. According to them, there are approximately 44 million OSN users in Pakistan who converse in Roman Urdu. They investigated the issue of cyberbullying using the Twitter platform, which allows users to interact in Roman Urdu. This is the first research in Roman Urdu that addresses cyberbullying behavior, to the best of the author's knowledge. They created a supervised machine-learning approach and suggested a lexicon-based model using a collection of characteristics taken from Twitter to address the problem. The findings show that the suggested lexicon-based technique is a viable option for identifying cyberbullying behavior in OSNs in Roman Urdu.

Rohit [14] on the multilingual cyberbullying detection system reveals that as the usage of social media has grown, so has the potential of its users to bully others. Because current methods for detecting cyberbullying are mostly focused on English-language writings, his thesis offers a novel method (dubbed the Multilingual Cyberbullying Detection System) for detecting cyberbullying in various languages (English, Hindi, and Marathi). To categorize the input data as bullying or non-bullying, it employs two techniques: machine learning-based and Lexicon-based. The goal of his study was to develop a distributed infrastructure for detecting cyberbullying in addition to detecting it. He created many prototypes (standalone, collaborative, and cloud-based) and tested them to identify cyberbullying on a variety of datasets in several languages. The results of the tests demonstrate that in all languages, the machine-learning model beats the lexicon-based approach. Also, Christiana et al. [15] on Multilingual Cyberbullying Detector (CD) Application for Nigerian Pidgin and Igbo Language Corpus stated that the sentiment analysis of Twitter data obtained with SNScrape demonstrate language-specific models that were flawless in identifying cyberbullying at reasonable runs.

III. METHODOLOGY

In empirical study, data gathering is critical. There are a few methods for data collection in research, all of which fall into two classes: primary and secondary sources of data [16]. Primary data, as the name suggests, is information that the specialist gathers interestingly, while secondary data is information that has effectively been assembled or made by others [17]. The data for this study came from both primary and secondary sources. The primary data was collected using a systematic closed-ended questionnaire, and the secondary data was gathered from the associated literature, which was reviewed by the researcher. A total of 60 structured questions were distributed tocyber security engineers in Nigeria with 5years and above practical expertise in cyber bullying detection and monitoring, versed in machine learning algorithms and all classifiers. For ease of accessibility, an online work marketplace for freelancing known as Upwork was used to categorize the professionals as regards their expertise and review rate; from which they were contacted across a spread offive (5) geo-political zones excluding Northeastern Nigeria. The questionnaire was created using information acquired from 15 similar publications via a literature study and content analysis. The surveys were graded on a 5-point Likert scale. The respondents were asked to rate how much they agreed or disagreed with the factors provided to them using this scale. For the objectives of this study, a random sample procedure was used. This strategy was chosen because it offered every one of the targeted respondents an equal chance of getting chosen [18].

The data were then calculated using the Statistical Package for the Social Sciences (SPSS) version 23. After that, we used principal component analysis and factor analysis to reduce the large set of variables we had to a small set that still contained most of the information in the original variables or large, resulting in dimensionality reduction, and then we used multiple linear regression of the factors. The survey results and some discussions are presented in the next section of the article.

IV. RESULTS

A. Respondents Multilingual Cyberbullying Detection Awareness and Usage 2019, 2020 and 2021

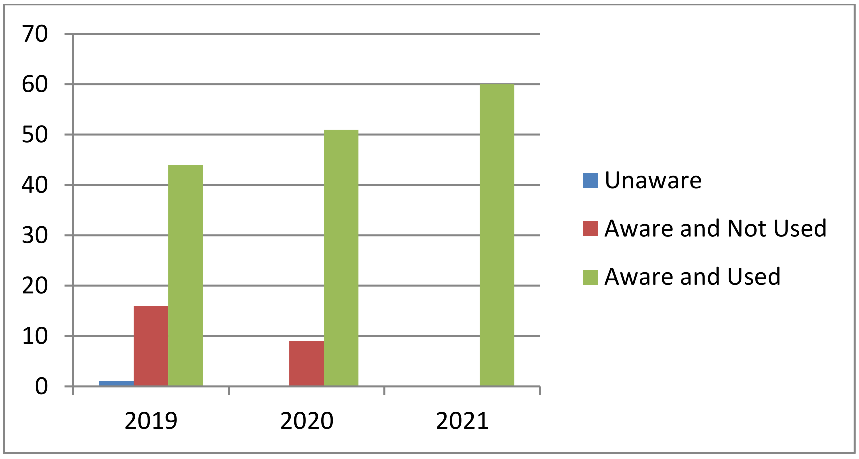

Fig I: Chart showing the Multilingual Cyberbullying Detection Awareness and Usage by Respondents for Three Years Span

Figure I show that, at first, the respondents, who represent a sample of all Nigerian industry professionals, were aware of the concept of multilingual cyberbullying detection; however, as the year progressed, the professionals did not only become aware of the concept but also used it for jobs in their fields of expertise. However, the lingering issue is: at what degree of multilingual cyberbullying detection maturity are they operating, and do they employ it? This study provides an answer to this question.

B. Variables Dimensionality Reduction

We used a multivariate analysis of Principal Component Analysis of Factor Analysis to reduce the big set of variables we had to a small set that still had much of the information in the original variables or a large set. We used it often for dimensionality-reduction of our data as an eigenvector-based multivariate analysis, concentrating on its three-to-four key areas: the determinant of the correlation matrix, KMO, and Bartlett's test, variance explanation with eigenvalue, and scree plot.

The test resulted in a determinant correlation matrix of 0.04, which is higher than the acceptable level of 0.001, indicating that the correlation matrix is not a non-positive definite (NPD), which means that none of the eigenvalues of the correlation matrix are non-positive numbers, implying that there are no linear dependencies among the variables and that there are no more variables in the analysis than there are cases.

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 10, Issue 3, pp: (7-15), Month: September - December 2023, Available at: www.noveltyjournals.com

Table I: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.614
Bartlett's Test of Sphericity	Approx. Chi-Square	962.574
	Df	300
	Sig.	.000

Table I presents two tests that recommend our data as appropriate for structure discovery. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy is a measurement that shows the amount of difference in our factors because of fundamental sources. High outcomes (around 1.0) recommend that factor analysis would be pertinent for our data. On the off chance that the value is under 0.50, the factor analysis results are probably not going to be especially useful. In any case, the KMO proportion of inspecting sufficiency of 0.614 in the table above demonstrates that the study's sample size is suitable and thus qualified for factor analysis.

Table II: Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.998	23.990	23.990	5.998	23.990	23.990	3.135	12.539	12.539
2	2.655	10.621	34.611	2.655	10.621	34.611	2.686	10.743	23.282
3	2.491	9.963	44.574	2.491	9.963	44.574	2.601	10.405	33.687
4	1.748	6.991	51.565	1.748	6.991	51.565	2.093	8.374	42.060
5	1.708	6.834	58.398	1.708	6.834	58.398	2.026	8.105	50.165
6	1.527	6.109	64.508	1.527	6.109	64.508	1.903	7.613	57.778
7	1.206	4.822	69.330	1.206	4.822	69.330	1.897	7.590	65.368
8	1.114	4.457	73.787	1.114	4.457	73.787	1.545	6.180	71.548
9	1.089	4.356	78.143	1.089	4.356	78.143	1.398	5.590	77.138
10	1.017	4.066	82.209	1.017	4.066	82.209	1.268	5.071	82.209
11	.780	3.119	85.329						
12	.598	2.390	87.719						
13	.584	2.335	90.054						
14	.476	1.906	91.959						
15	.421	1.685	93.645						
16	.349	1.395	95.040						
17	.298	1.193	96.233						
18	.255	1.019	97.253						
19	.190	.759	98.011						
20	.158	.632	98.643						
21	.137	.549	99.192						
22	.105	.420	99.612						
23	.046	.182	99.794						
24	.037	.149	99.943						
25	.014	.057	100.000						

Extraction Method: Principal Component Analysis.

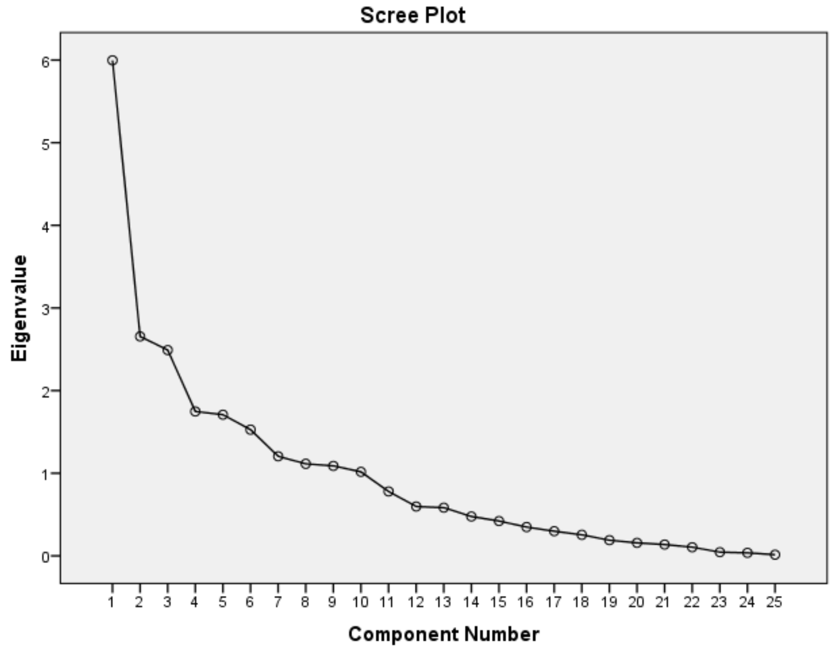

Fig. II: Scree Plot showing the eigenvalue of factors reduced to ten from twenty-five using eigenvalue set at 1 as benchmark.

From Table II and Figure II, the obtained factors were reduced to ten (10) factors from the original twenty-five (25) factors. Since the eigenvalue was set at 1 as the benchmark, all factors with eigenvalues from the benchmark and above were considered not multi-colinear and will serve as the independent variables of the study. The scree plot further x-rayed this outcome; therefore, the factors considered by results are listed below with bold characters.

Table III: Component Matrix[a]

	Component									
	1	2	3	4	5	6	7	8	9	10
1. Language ambiguity	**.175**	**.198**	**.367**	**-.251**	**-.494**	**-.442**	**.006**	**.133**	**.311**	**-.002**
2. Sarcasm dialects classification	**.681**	**.022**	**.259**	**-.310**	**-.336**	**-.266**	**.041**	**-.009**	**.115**	**.053**
3. Data scarcity	**.618**	**-.153**	**-.164**	**.119**	**.383**	**-.255**	**.169**	**-.228**	**-.129**	**.378**
4. Non-trusted data source	**.025**	**-.109**	**-.133**	**.263**	**.593**	**-.517**	**.041**	**.122**	**.153**	**.261**
5. Moribund slang trend	**.021**	**.218**	**.366**	**.169**	**-.005**	**.571**	**.464**	**-.075**	**.291**	**.019**
6. costly collection and annotation of training data	**-.686**	**.154**	**.030**	**.397**	**.159**	**.026**	**.085**	**.013**	**.319**	**.072**
7. Feature learning challenge	**.589**	**-.213**	**-.160**	**-.380**	**.022**	**-.055**	**-.173**	**.119**	**.187**	**.040**
8. Difficulty in developing novel classifier for multilingual text detection.	**.334**	**-.015**	**.218**	**-.133**	**-.266**	**.191**	**.514**	**.230**	**-.081**	**.264**
9. Domain influence	**.395**	**-.263**	**-.262**	**.033**	**.294**	**.189**	**.210**	**.189**	**.007**	**-.591**
10. Languages characters inconsistent heights, strokes, and writing format.	**.630**	**-.342**	**-.161**	**.328**	**-.205**	**.085**	**.082**	**.232**	**.031**	**.044**
11. Restrictions on data access	.445	-.067	-.212	.667	-.251	.067	.025	.257	.016	.054
12. Small Datasets	.364	.809	-.081	.145	.110	.110	-.162	-.126	.229	-.054

ISSN 2394-7314

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 10, Issue 3, pp: (7-15), Month: September - December 2023, Available at: www.noveltyjournals.com

13. Non-generalizable linguistic features	.099	.797	-.231	-.118	.019	-.023	.064	.235	-.305	.048
14. Difficulty in discerning emotions	.464	-.072	-.453	-.083	.029	.284	-.108	.070	.258	.355
15. Complexity of classifying vernacular invective posts	.268	-.130	-.613	-.258	-.022	.242	-.273	.348	.257	.002
16. Non- consideration of distributional semantic features	.575	-.143	-.223	-.311	.199	.119	.082	-.352	-.121	-.234
17. Limited database for Igbo, Hausa and Yoruba languages	.553	-.036	-.014	.548	-.336	-.129	-.197	-.161	-.136	-.173
	.795	-.097	.054	.246	-.256	-.039	-.114	-.311	-.053	-.030
18. Possibility of OSN ban	.503	.651	-.041	.057	.089	.045	-.188	-.235	.292	-.061
19. Preprocessed online social network conversation test challenge	.252	.661	-.165	.001	.035	.018	-.019	.289	-.465	.051
20. Frequency Extraction challenge	.768	.020	-.145	-.128	.045	-.020	.345	-.206	-.102	.083
21. Experts inadequacy	.384	.224	.008	.000	.242	-.448	.389	.171	.263	-.327
22. Software availability Issue	.383	-.120	.541	.119	.181	-.115	-.188	.365	-.090	-.149
23. Lack of collaboration among experts	.501	-.062	.659	-.042	.383	.225	-.182	.090	-.074	.081
24. Funding challenge	.539	-.028	.653	-.023	.290	.253	-.235	.069	.014	.078
25. Tribal differences	.421	-.021	.641	-.019	.271	.261	-.241	.052	.009	.065

Extraction Method: Principal Component Analysis.
a. 10 components extracted.

C. Critical Barriers to BIM Adoption in Nigeria

Table IV: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig.F Change
1	.953	.909	.900	.64550	.909	107.728	9	50	.000

Table IV shows that the Multiple Regression Coefficient "R" is 0.953, indicating that the Dependent variable Y and the Independent variables have a strong positive association (X1-X10). R2, or the multiple determination coefficient, was found to be 90.9 percent.

Table V below shows the F-Test (ANOVA) table.

Table V: ANOVA[a]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	224.433	9	44.887	107.728	.000[b]
	Residual	22.500	50	.417		
	Total	246.933	59			

The outcome of the ANOVA analysis table shown above indicates that it is possible to assess if there is a statistically significant link between the variables. The significance level is 0.000, which is less than 0.05, as shown in Table V. This indicates that the variables have a statistically significant association.

Table VI: Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	-8.632	3.871		-2.230	.030
	Language ambiguity	.397	.077	.335	5.133	**.000**
	Sarcasm dialects classification	.384	.085	.321	4.934	**.000**
	Data scarcity	.357	.069	.302	4.521	**.000**
	Non-trusted data source	.368	.066	.686	2.542	**.010**
	Moribund slang trend	-.059	.058	-.129	-1.006	.319
	Costly collection and annotation of training data	.221	.051	.210	4.362	**.000**
	Feature learning challenge	-.067	.069	-.141	-1.211	.332
	Difficulty in developing novel classifier for multilingual text detection.	.362	.111	.165	2.901	**.000**
	Domain influence	.367	.092	.317	4.793	**.000**
	Languages characters inconsistent heights, strokes, and writing format.	.353	.128	.177	2.754	**.008**

Table VI shows the coefficients of each of the explanatory variables X1-X10, which are found to be 0.000, 0.000, 0.000, 0.010, 0.319, 0.000, 0.332, 0.000, 0.000, 0.000, 0.000, 0.000, and 0.008 respectively. This result is further explained in ANOVA Table 9. It reveals that X1 (Language ambiguity), X2 (Sarcasm dialects classification), X3 (Data scarcity), X4(Non-trusted data source), X6 (Costly collection and annotation of training data), X8 (Difficulty in developing novel classifier for multilingual text detection), X9 (Domain influence) and X10 (Languages characters inconsistent heights, strokes, and writing format.) with the significance level 0.000, 0.000, 0.000, 0.010, 0.000, 0.000, 0.000 and 0.008 respectively are all variables that are responsible for that significance of the model, and thus are the critical technological factors affecting multi-lingual cyberbullying detection in Nigeria.

Hence, the model is thus summarized in the equation below.

Y=-8.632+0.397X1+0.384X2+0.357X3+0.368X4+-0.059X5+0.221X6+0.067X7+0.362X8+0.367X9+0.353X10------------------------------- (1)

From the analysis of the study, the explained variable, which is the detection of multilingual cyberbullying (Y) was shown to have some levels of association with the explanatory factors. The explanatory factors and multilingual cyberbullying detection (Y) have a modest positive connection (R = 0.953) (X1-X10). The coefficient of determination was also determined to be 90.9 percent. The explanatory factors in this statistic explain the variance in multilingual cyberbullying detection (Y). This means that the explanatory factors are unable to account for 9.9% of the variance in multilingual cyberbullying detection (Y), which might be referred to as model error. The modified coefficient of determination value of 90.0 percent was also found, indicating that the explanatory factors account for 10% of the variation in multilingual cyberbullying detection (Y). As a result, the F-statistics with a value of 107.728 and a likelihood of 0.000 show that the independent factors are jointly important in explaining the variance in the dependent variable, multilingual cyberbullying detection (Y).

Individual contributions of the independent variables are examined in the model, X1-X4, X6, X8-X10 showed a positive significant relationship with multi-lingual cyberbullying detection (Y), depicting that a 1% increase in X1-X4, X6, X8-X10 will cause a 0.397%, 0.384%, 0.357%, 0.368%, 0.221%, 0.362%, 0.367%, or 0.353% increase in Y, whereas X5 and X7 are not significant.

Table VII: Ranking of the factors according to their coefficient or impact significance.

Variable Designation	Technological factors	Rank
X1	Language ambiguity	1st
X2	Sarcasm dialects classification	2nd
X9	Domain influence	3rd
X3	Data scarcity	4th
X8	Difficulty in developing novel classifier for multilingual text detection	5th
X6	Costly collection and annotation of training data	6th
X10	Languages characters inconsistent heights, strokes, and writing format	7th
X4	Non-trusted data source	8th

V. CONCLUSION AND FUTURE WORKS

This paper has provided an empirical evaluation of the critical technological factors affecting multilingual cyberbullying detection in Nigeria. The results of the study clearly depict that the core technological variables impeding cyberbullying detection of both the English language and any of the Nigerian primary indigenous languages of Hausa, Yoruba, and Igbo are language ambiguity, which stands for the challenge of logical semantics, stylistic equivalences, and slang contextual meaning; sarcasm dialect classification, source and target domain influence to deal with unlabeled data, data scarcity due to the non-availability of a database for the indigenous languages; and non-trusted sources of data for the languages, such as linguists in the various indigenous languages. There is difficulty in developing novel classifiers for multilingual text detection, costly collection, and annotation of training data and language characters with inconsistent heights, strokes, and writing formats. However, based on verifiable research, no study has been carried out on multilingual cyberbullying detection comprising any of Nigeria's indigenous languages (Hausa, Yoruba, Igbo, Efik, etc.), and as such, it is recommended that research be conducted in that direction so as to capture latent cyberbullying activities that cannot be detected in English languages because, on a daily basis, cyberbullying is carried out in indigenous languages and their connecting slangs

REFERENCES

[1] Margare Jackson., Wanda Cassidy and Karen N. Brown, “You were born ugly, and you’ll die ugly too. Cyber-bullying as relational aggression,” Education Journal: Special Issue on Technology and Social Media, vol. 15, no 2, 2019. Available:https://www.academia.edu/21601182/Sticks_and_Stones_Can_Break_My_Bones_But_How_Can_Pixels_Hurt_Me_Students_Experiences_with_Cyber_Bullying

[2] Jaimee Stuart, Riley Scott, Calvin Smith and Molly Speechley, “Parents' anticipated responses to children's cyberbullying experiences; action, education and emotion”, Children and Youth Services Review, vol. 136, 2022. Available: https://www.sciencedirect.com/science/article/abs/pii/S0190740922000342

[3] Constanza Moretti and Damián Herkovits, “Victims, perpetrators, and bystanders: a meta-ethnography of roles in cyberbullying”, Cadernos de Saúde Pública, vol. 37, 2021. Available: https://www.scielo.br/j/csp/a/5tqGgm7fGVs8xDsKqmW9v7r/abstract/?lang=en

[4] Siswanta Maya Sekarwangi and Aris Triharyanto, “The phenomenon of cyberbullying in the virtual social media communication community”, In 2nd Annual Conference on blended learning, educational technology and Innovation (ACBLETI 2020). New York, Atlantis Press, pp. 409-414, 2021. Available: https://www.atlantis-press.com/proceedings/acbleti-20/125957859

[5] Adeola O. Opesade and Abiodun O. Adetona, “An assessment of internet use and cyber-risk prevalence among students in selected Nigerian secondary schools”, Journal of Cybersecurity Education, Research and Practice, vol. 2020, no 2, pp.3-10, 2021. Available: https://digitalcommons.kennesaw.edu/jcerp/vol2020/iss2/3/

[6] Ochola Ochoche Ijachi, "Social media access and cyberbullying—a Nigerian Perspective", International Journal of Innovative Studies in Medical Sciences, vol. 3, no. 3, pp.5-9, 2019. Available: https://ijisms.com/storage/Volume3/Issue3/IJISMS-030302.pdf

[7] Adejoke O. Adediran, "Cyberbullying in Nigeria: Examining the adequacy of legal responses", International Journal for the Semiotics of Law, vol. 34, no. 4, pp. 965-984, 2021. Available: https://link.springer.com/article/10.1007/s11196-020-09697-7

[8] Whittaker E. and Robin M.K., "Cyberbullying via social media", Journal of School Violence, vol.14, no. 1, pp.11-29, 2015. Available: https://www.tandfonline.com/doi/abs/10.1080/15388220.2014.949377

[9] Jacobs Gilles, Cynthia Van-Hee and Veronique Hoste, "Automatic classification of participant roles in cyberbullying: Can we detect victims, bullies, and bystanders in social media text?", Natural Language Engineering, vol. 28, no. 2, pp.141-166, 2022. Available: https://www.cambridge.org/core/journals/natural-language-engineering/article/abs/automatic-classification-of-participant-roles-in-cyberbullying-can-we-detect-victims-bullies-and-bystanders-in-social-media-text/A2079C2C738C29428E666810B8903342

[10] Chuka Fred Ononye and Nkechinyere Juliana Nwachukwu, "Metalinguistic evaluators and pragmatic strategies in selected hate-inducing speeches in Nigeria", Indonesian Journal of Applied Linguistics, vol. 9, no. 1, pp. 48-57, 2019. Available: https://pdfs.semanticscholar.org/0447/740c3ff8d95270f16c76437a3347d5939bb0.pdf

[11] Nairaland, "Yoruba people and insult,"2019. [Online]. Available: https://www.nairaland. com/4341512/yoruba-demon-joke-tribal-insult[Accessed 10 December 2021].

[12] Rounak Ghosh, Siddhartha Nowal and Manju G., "Social media cyberbullying detection using machine learning in Bengali language," International Journal of Engineering Research & Technology (IJERT), vol. 10, no 5, pp. 190-193, 2020. Availabe: https://www.ijert.org/social-media-cyberbullying-detection-using-machine-learning-in-bengali-language

[13] Kazim Raza Talpur, Siti Sophiayati Yuhaniz, Nnilam Nur Binti Sjarif and Bandeh Ali Amir, "Cyberbullying detection in Roman Urdu language using lexicon based approach," Journal of Critical Reviews, vol.7, no 16, pp, 834-848, 2020. Available: https://www.researchgate.net/publication/341575259_Cyberbullying_Detection_in_Roman_Urdu_Language_Using_Lexicon_Based_Approach#fullTextFileContent

[14] Rohit S.Pawar., "Multilingual cyberbullying detection system," MSc. Thesis, Department of Computer Science, Purdue University, 2019. Available: https://scholarworks.iupui.edu/bitstream/handle/1805/18942/MULTILINGUAL%20CYBERBULLYING%20DETECTION%20SYSTEM.pdf?sequence=1&isAllowed=y

[15] Christiana A. Okoloegbo, Udoka F. Eze, Gloria A. Chukwudebe, and Obi C. Nwokonkwo. "Multilingual Cyberbullying Detector (CD) Application for Nigerian Pidgin and Igbo Language Corpus," in Proceedings of the 2022 5th Information Technology for Education and Development (ITED), 01-03 November 2022, Abuja, Nigeria, 2022, doi: 10.1109/ITED56637.2022.10051345. pp. 1-6, Published by IEEE, Available: https://ieeexplore.ieee.org/abstract/document/10051345

[16] Douglas M., "Sources of data," 2015. [Online]. Available: http://www.online etymology dictionary/data [Accessed 12 December 2021].

[17] Mesly Olivier., "Creating models in psychological research," Etats-Unis: Springer Press, 2015. Available: https://books.google.com.ng/books?hl=en&lr=&id=11ayBwAAQBAJ&oi=fnd&pg=PP7&dq=Mesly+Olivier

[18] Gary J. Burkholder, Kimberley A. Cox., Linda M. Crawford. and John H. Hitchcock., "**Research design and methods: An applied guide for the scholar-practitioner,**" New York, Sage Publishers, 2019. Available: https://www.abebooks.com/9781544342382/Research-Design-Methods-Applied-Guide-1544342381/plp