Fake News Detection in News Articles written in English, Filipino, and Spanish using Machine Learning Algorithm LinearSVC

Ramilo V. Orejana Jr.
College of Science and Computer Studies
De La Salle University - Dasmariñas
orv1728@dlsud.edu.ph

ABSTRACT
Misinformation spreads quickly in social media through the various news articles shared on online platforms. The speed and scale of its spread is not achievable to be controlled solely by authorized agencies, as shown by the negative effects it has on decision making by the public. To help with its detection, the study aims to create a Machine Learning model that can classify news articles into either fake news or real news. The created system used a LinearSVC model alongside Direct Quotes Removal data cleaning and TF-IDF feature extraction. The classifier system was able to successfully classify between fake news and real news articles, achieving an accuracy of 91.1%.

Keywords
Fake News Detection, Spanish News, English News, Filipino News, Machine Learning, LinearSVC

1. INTRODUCTION
Misinformation can be found in social media and other platforms people use to communicate, and affects their decisions and outlook on certain topics. For instance, [1] found that COVID-19 vaccine misinformation was responsible for 27% of vaccine hesitancy-related comments found in Filipino TikTok videos.

[2] states that part of the reason why modern disinformation in social media is dangerous is because it panders to preconceived notions and emotions to convince people to believe the misinformation. This is corroborated by [3] where it was found that misinformation usually uses pathos, a literary technique that uses deliberate word choice to appeal to human emotion.

[4] states that social media has played a big role in the spread of misinformation today. It emphasizes that this is because notable users in the form of elderly and young adolescents are vulnerable because of their lack of media literacy. However, this does not mean misinformation is limited to social media and regular citizens. As [5] shows, the Philippine Department of Health has implicitly declared vape products as non-hazards by passing monitoring duties to the Department of Trade and Industry, despite a lack of reliable studies that consider vaping products as safe, partially because of vape industries' marketing propaganda.

The study aims to contribute to the field of Fake News Detection in the following goals:

1. Attempt to create a machine learning classifier model that can classify in three national languages.
2. Test newly proposed data cleaning technique for news articles called Direct Quotes Removal
3. Observe the performance of LinearSVC in Fake News Detection

To combat misinformation spread, the study created a classifier model that can identify if a news article is likely to be "fake news" containing misinformation, or real news. Three languages were chosen to be used in training the model, to observe the viability of using multilingual fake news detection models. English, Spanish, and Filipino were chosen. English and Spanish are two languages that are generally high resource languages in the field of NLP. However, for the niche of Fake News Detection, Spanish and Filipino are both low resource languages. Thus, these two languages were chosen alongside English to bring more attention to them in the NLP Fake News Detection field.

Additionally, after analysis of common data preprocessing techniques, the study has found a lack of data processing geared towards word choice in news articles. Specifically, studies that use feature extraction techniques similar to Bag-of-Words, which use word frequencies, do not consider Direct Quotes, which contain words that were not specifically chosen by the authors of the article. Thus, the study will also test a new data cleaning technique for news articles called Direct Quotes Removal.

2. RELATED WORKS
2.1 Datasets Used in Other Studies
[6] created an annotated Fake News Corpus in the Filipino Language. They gathered news articles from various news articles publishing websites. Real news articles were gathered from reputable news sites like *Manila Times* and *Philippine Daily Inquirer*, while misinformation-containing fake news articles were gathered from lists of sites verified to contain fake news by organizations such as the Senate and CBCP [7]. [8] used a similar approach using the same list of websites to create an English Fake News Corpus. [9] created a Spanish Fake News Corpus. Alongside the site credibility approach that the earlier two datasets used to annotate the articles as fake or real, the Spanish news articles were also manually verified for credibility, either by the corpus creators or by trusted Spanish news verifier sites such as *VerificadoMX*.

2.2 Fake News Detection Methods Used in Other Studies
2.2.1 Multilingual Models
[10] constructed a sentiment analysis model. The dataset used in the study used Facebook comments as an alternative source, more specifically the comments were sourced from e-commerce posts.

The dataset was comprised of 1000 comments that were written in pure Filipino, pure English, and mixed, also known as Taglish. The polarity of the data was split between positive comments and negative comments. To preprocess the data, stop words were removed and stemming was also applied. To classify, the data was fed into a Naïve Bayes model which achieved 77% accuracy.

Similarly [11] constructed Bilingual models to detect online fake news, and concluded that Linear Support Vector Classifier or LinearSVC performed the best, with an accuracy of 93.29% and an F1 score of 0.93. Their dataset was constructed using the English dataset called ISOT fake news with more than 12600 credible articles as well as 12600 fake news articles, merged with a Bengali dataset, which they trimmed to 7000 credible articles and 1000 fake news articles. The data was preprocessed by tokenization, removing stop words with their own constructed list of Bengali stop words, then punctuation and other characters removed, after which the tokens underwent stemming. Features were then extracted by converting the text into numerical format using TF-IDF, where its performance was enhanced by using N-Grams. By using bigrams, they discovered that they obtained the highest accuracy with the LinearSVC, while using trigrams gave the highest F1 score, also with LinearSVC.

2.2.2 Models that used TF-IDF and/or LinearSVC

TF-IDF holds deeper linguistic context than Count Vectorizer. These are factors such as considering few occurrences as insignificant because they are not used in the document frequently enough, as well as excessive occurrences for insignificant words such as "said". It is also suitable for the deliberate word choice that [2, 3] shows is correlated with misinformation. It will be discussed further in the methodology of the study. Looking at the success rate of models that used TF-IDF:

[12] constructed a dataset containing 35,027 real news articles and 37, 106 fake news articles for a total of 72, 134 articles. The dataset was constructed from multiple other datasets, including ones taken from sources such as Kaggle, as well as gathered from sites like Reuters and Buzzfeed Political. They used it in a study where 7 machine learning algorithms were used, including their own constructed algorithm called WELFake, which uses Word Embedding over Linguistic Features for text classification. They then calculated Linguistic features such as the writing pattern category which finds the numbers of symbols and number of capital letters, psycholinguistics which is concerned about text polarity and subjectivity, etc. A total of 20 of these features were selected. For feature extraction, term frequency-inverse document frequency (TF-IDF) was used. The study concluded that using their own constructed WELFake model, they were able to achieve a 96.37% accuracy. They also found that by analyzing the other 6 machine learning methods in terms of evaluation metrics such as accuracy, precision, recall, and F1-score, and found that the SVM obtained the most accurate results.

[13] performed a comparative study consisting of multiple machine learning models trained on classifying Fake News about COVID-19. The machine learning algorithms consisted of 12 algorithms, including Support Vector Machine, Logistic Regression, Naïve Bayes, etc. The dataset used top train the model was named "COVID Fake News Dataset", published by Sumi Banik on *Coronavirus Disease Research Community-Covid-19*. It contains 10000+ articles labelled as real and fake, where the content was the article headline as well as the article content. The data was preprocessed by removing empty words, short words that only contained 3 characters, removing links, etc. The data was then vectorized by converting the text to numerical data in a TF-IDF format. The data was fed to the different models, and their performances in the evaluation metrics of precision, recall, and F1 score were compared. It was concluded that the Convolutional Neural Network and BiLSTM (bidirectional Long Short Term Memory) models had the best performance with 97% accuracy.

[14] also underwent a comparative study of machine learning algorithms to compare their performance in Fake News Detection. They used six algorithms, consisting of Decision Tree, Random Forest, Support Vector Machine, Naïve Bayes, KNN (k-Nearest Neighbors), and XGboost. They used the publicly availabled LIAR-PLUS dataset, which they preprocessed by removing punctuation marks and stemming such as removing suffixes. Feature extraction was then done by choosing linguistic features such as word length, then n-gram features were extracted using TF-IDF Vectorizer. The dataset was split between 70% for training and 30% for testing.; The clean data was then fed to the different classification models, which were evaluated using accuracy, precision, recall, and f1-score. Their results showed that they achieved their highest accuracy of 75% with the XGboost model, while the next highest accuracy models were SVM and Random Forest with 73% accuracy.

Looking at the results obtained in this study, it can be seen that TF-IDF usually produces accuracy results of 70-90% in Fake News Classification. It is also a trend to see LinearSVC and other SVMs perform well in the classification task. Thus, both are viable tools to use in the creation of the classifier model.

3. METHODOLOGY

3.1 Datasets Used

This study used the earlier discussed previously gathered datasets, composing of news articles labelled as fake or real from the languages English, Filipino, and Spanish. It selected these datasets from previous studies that were verified to be reliable as the articles are published and peer reviewed. These datasets are annotated, with labels identifying whether the specific row or article contains misinformation or not.

Table 1. Datasets Used in the Study

Dataset Name and Language	Name of Study/Competition where Dataset was used	Reason for Selection/Reliability	Link to GitHub repository of dataset	Size of Dataset
Philippine Fake News Corpus (English)	Computing the Linguistic-Based Cues of Fake News in the Philippines Towards its Detection	Published in the "Proceedings of the 9th International Conference on Web Intelligence, Mining and Semantics"	https://github.com/aaroncarlfernandez/Philippine-Fake-News-Corpus	14, 802 real news, and 7,656 fake news articles, for a total of 22,458 articles

Fake News Filipino Dataset (Filipino)	Localization of Fake News Detection via Multitask Transfer Learning	Published in the "LREC 2020 Proceedings"	https://github.com/jcblaisecruz02/Tagalog-fake-news/tree/master	3206 news articles, exactly half of which are real and exactly half are fake.
The Spanish Fake News Corpus Version 1.0 (Spanish)	Overview of MEX-A3T at IberLEF 2020: Fake News and Aggressiveness Analysis in Mexican Spanish	Published in the IberLEF 2020: Iberian Languages Evaluation Forum	https://github.com/jpposadas/FakeNewsCorpusSpanish	491 real news, 480 fake news, for a total of 971 articles
The Spanish Fake News Corpus Version 2.0 (Spanish)	Overview of FakeDeS at IberLEF 2021: Fake News Detection in Spanish Shared Task	Published in the IberLEF 2021: Iberian Languages Evaluation Forum	https://github.com/jpposadas/FakeNewsCorpusSpanish	286 real news, 286 fake news, for a total of 572 articles

3.2 Dataset Cleaning and Preprocessing

Minimal dataset cleaning was used to verify the data to preserve its state as much as possible. All rows with blank columns or metadata were dropped, and the labels were standardized to "Credible" for real news and "Not Credible" for fake news. The Spanish dataset now had 1468 entries, the Filipino dataset had 3206, and the English dataset had 13317. To prevent bias towards English, a minimal amount of 5500 cleaned English articles was used, for a total of 10,174 articles across the three languages. After this, each language's dataset was split into 70% training and 30% testing data.

The study theorizes that direct quotes reduce the accuracy of a Bag-Of-Words model, because the words included in the quotes are selected by the speaker, not by the journalist or other writers of the news article. This harms fake news detection where it has been established that deliberate word choice in articles can determine whether it contains misinformation, as the deliberate word choice of the author/s could be covered by the quoted words.

Thus, to further clean the dataset, Direct Quotes Removal is performed. To accomplish this, remove_quotes was created, a custom-made function made by the proponents of the paper that utilizes the ".sub()" function used in the regex Python library. It eliminates direct quotes that are potentially harmful to the accuracy of the dataset. This function uses regular expressions to match direct quotes by looking for a set of two double quotation marks (") as character delimiters on both ends, as well as the text between them, as this is a standard for direct quotes (APA, 2022). It then substitutes the detected direct quote and substitutes it with a whitespace.

3.3 Feature Extraction

Dataset extraction was performed using the Bag of Words technique of TF-IDF, which stands for Term Frequency – Inverse Document Frequency. This technique gauges how important a particular word or term is to its class. The study performed this technique using the following:

3.3.1 Count Vectorizer

	000	10	100	17	18th	1982	2012	2015	2016	2018	...	worth	would	wounded	wrong
0	0	0	0	0	0	0	0	1	0	0	...	0	0	0	4
1	0	0	0	0	0	0	1	0	0	0	...	0	2	0	0
2	0	0	0	0	0	2	0	0	0	0	...	0	4	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
4	0	0	0	1	0	0	0	0	0	2	...	1	0	1	0
5	0	0	0	0	2	0	0	0	0	0	...	0	0	0	0
6	4	1	3	0	0	0	1	0	1	0	...	0	0	0	0

Figure 1. Count Vectorized Sample Data

The Count Vectorizer algorithm takes text input and returns a numerical matrix representation of the number of times a particular word has appeared in its document. The sci-kit learn library's implementation of Count Vectorizer has a built-in tokenizer, lowercase, and punctuation remover. The implementation used by this paper is the scikit-learn machine learning library for Python. The study used the default parameters for this application.

3.3.2 TF-IDF Transformer

	000	10	100	17	18th	1982	2012	2015	2016	2018	...	worth	would	wounded	wrong
0	0.00000	0.00000	0.00000	0.000000	0.000000	0.000000	0.000000	0.069917	0.00000	0.000000	...	0.000000	0.000000	0.000000	0.27967
1	0.00000	0.00000	0.00000	0.000000	0.000000	0.000000	0.027452	0.000000	0.00000	0.000000	...	0.000000	0.054904	0.000000	0.00000
2	0.00000	0.00000	0.00000	0.000000	0.000000	0.050675	0.000000	0.000000	0.00000	0.000000	...	0.000000	0.084129	0.000000	0.00000
3	0.00000	0.00000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	...	0.000000	0.000000	0.000000	0.00000
4	0.00000	0.00000	0.00000	0.047173	0.000000	0.000000	0.000000	0.000000	0.00000	0.094345	...	0.047173	0.000000	0.047173	0.00000
5	0.00000	0.00000	0.00000	0.000000	0.135302	0.000000	0.000000	0.000000	0.00000	0.000000	...	0.000000	0.000000	0.000000	0.00000
6	0.09964	0.02491	0.07473	0.000000	0.000000	0.000000	0.020678	0.000000	0.02491	0.000000	...	0.000000	0.000000	0.000000	0.00000

Figure 2. TF-IDF Sample Data

On its own, Count Vectorizer contains only shallow linguistic meaning and cannot be reliably used as a feature extractor for text. Thus, the extracted data by Count Vectorizer is retransformed using TF-IDF Transformer. This is because TF-IDF holds deeper linguistic context than Count Vectorizer. These are factors such as taking into account few occurrences as insignificant because they are not used in the document frequently enough, as well as excessive occurrences for insignificant words such as "said".

The TF-IDF Transformer takes the matrix the Count Vectorizer has made and converts it to the TF-IDF values, where:

- TF = Term Frequency; the number of times a word token has occurred in the document over the number of words in the document.

  Term Frequency (TF) uses the formula $TF(t, d) = t_n / d_n$, where:

  t_n = the number of times a specific term has appeared in a document, and;

  d_n = the total amount of words in the document.

- IDF = Inverse Document Frequency; log of the ratio between the number of documents in the corpus and the number of documents the word is in.
  IDF is obtained by getting the logarithm of the number of documents in the corpus divided by the number of documents where the term occurs, or:

  $IDF = \log(C_n / C_{n(t)})$ where:
  C_n = number of documents in the corpus
  $C_{n(t)}$ = number of documents in the corpus where the term occurs

- TF-IDF = The values of TF and IDF are multiplied together.

The implementation used by this paper is the scikit-learn machine learning library for Python. The study used the default parameters for this application.

3.4 Model Training and Evaluation

3.4.1 LinearSVC

A Support Vector Machine implementation from the sci-kit learn library. It finds a hyperplane that most cleanly separates the classes with the largest reasonable margin while trying to avoid overfitting by allowing some misclassifications. In this study, two linear SVC models were used, one trained on the dataset with Direct Quotes Removal, and another trained with no Direct Quotes Removal to serve as a baseline performance.

This study uses the implementation of the sklearn library, where the LinearSVC model is similar to the base SVC model using a linear kernel, with slight differences for optimization. The model was not fine-tuned and only the default parameters were used in the study.

3.4.2 Performance Metrics

This study uses the evaluation metrics of accuracy, precision, and recall to evaluate the model's performance. All three were obtained using the relevant functions imported from the sklearn library.

4. RESULTS AND DISCUSSION

In this section the performance of the model is recorded according to the evaluation metrics of accuracy score, precision score, and recall score. Each metric was measured on 8 different datasets. The datasets were composed of two groups, one with Direct Quotes Removal (DQR) performed, and the other with DQR not performed, which henceforth will be referred to by the paper as the "plain" dataset. Each is broken down further into four subgroups, consisting of the three languages separated into their individual testing datasets, and one overall dataset consisting of the combined language datasets.

4.1 Results

Table 2. Results Obtained from Different Models

Testing Results				
		Accuracy	Precision	Recall
LinearSVC Model trained on plain datasets	**Combined Languages**	**0.911**	**0.904**	**0.929**
	English	0.943	0.951	0.944
	Filipino	0.926	0.928	0.923
	Spanish	0.757	0.842	0.731
LinearSVC Model trained on datasets that underwent Direct Quotes Removal (DQR)	**Combined Languages**	**0.915**	**0.905**	**0.936**
	English	0.946	0.955	0.946
	Filipino	0.932	0.945	0.920
	Spanish	0.761	0.842	0.736

4.2 Interpretation of Results

4.2.1 Interpretation of Results

The LinearSVC algorithm was overall successful in classifying between real or fake datasets, achieving greater than 90% results in all three of accuracy, precision, and recall in the combined languages dataset. Additionally, for all 3 languages as well as the combined language, the model trained on data that underwent Direct Quotes Removal increased on almost all three evaluation metrics. However, the increase in performance cannot be declared as statistically significant without a statistical analysis, because of the small amount of improvement. A statistical analysis can be done to confirm the significance of the improvements, but this is outside the scope of the study.

4.2.2 Filipino Results

The Filipino results are notable because it is the only language where Direct Quotes Removal resulted in a lower performance, where the recall score of the dataset that did not undergo DRQ had 0.923, a higher score than the dataset that underwent DQR, which had 0.920. However, the difference is not large enough to simply declare it significant. Additionally, it is also the only language where DRQ could be simply said to have a significant effect. The precision score of the dataset that underwent DRQ was 0.945, a higher score than the dataset that did not undergo DRQ, which had 0.928, a difference of 0.17 or 1.7% improvement. However, the difference is still not large enough to simply declare it significant without a statistical analysis performed.

4.2.3 Spanish Results

The Spanish results are notable in that while they are high enough to feasibly declare that the models can classify the "realness" of the news articles, they still achieved particularly low results among the three languages. This may be because of the low dataset size of the language in comparison to the other two languages.

It is also notable that in both models with or without DQR, its precision results are significantly higher than its accuracy and recall scores. This means that compared to its other abilities, its ability to avoid classifying a fake news article as real is higher. Low as the accuracy may be, this may be better as in practical terms, being able to avoid misinformation is the focus of the study, and classifying a real news article as fake is not nearly as damaging as classifying misinformation as real.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

After analyzing the performance of the model, we can conclude that both models trained with and without Direct Quotes removal, using TF-IDF as a feature extraction technique was successful in creating a classifier of Fake and Real News articles. However contrary to what was first hypothesized, the Direct Quotes Removal technique was not able to produce a significant difference in the performance of the models.

5.2 Recommendations

According to the results, the models in the study were not able to classify the Spanish news articles very successfully in comparison to the other languages analyzed in the study. When analyzing the difference between the Spanish dataset and the other languages, the main difference in their size is what stands out, especially because the Spanish dataset alone can be considered even when not relating it to other datasets. Therefore, the researcher recommends that more data gathering is done for Spanish news articles, which may solve the accuracy problem of the model.

In this study, TF-IDF was used as a feature extraction format that has seen success in many different studies. However, using feature extraction in the style that this study used creates a Bag-Of-Words format, which does not contain the context that the words were used in. The researcher recommends that techniques that preserve the context of the news articles be tried to improve the accuracy of the model, such as using n-grams.

Lastly, the Direct Quotes Removal implementation in the study is admittedly lacking. One of the challenges that can be faced in using Regular Expressions is the presence of multiple formats of characters, such as unicode and ASCII. In the implementation of this study, the detection of direct quotes relied on the character of neutral double quotation marks ("). Notably however, different journals may use different styles such as using left double quotation marks(“) and right double quotation marks (”) to denote a direct quote. Unfortunately, regular expressions struggle to detect these kinds of double quotation marks because of their existence in different formats, thus regular expressions is not able to detect al direct quotes without an exhaustive list of formats of double quotation marks. Additionally, some journals may choose to use single quotation marks (‘’) to denote quotes. Using regular expressions to detect these quotes is dangerously unreliable as it may conflict with apostrophes used in various words such as possessive nouns. Lastly, human error in the usage of quotation marks by the writer of the article may lead to lacking or excessive detection of direct quotes.

All of these factors mean that there is a big need to improve the Direct Quotes Removal implementation. This may be the reason why Direct Quotes Removal did not have a significant effect on the performance of the models. Thus, the researcher recommends that more research is done to improve the implementation of Direct Quotes Removal in other studies to see exactly how this technique affects model classification performance.

6. REFERENCES

[1] Berdida, D.J.E. et al. 2022. Filipinos' COVID-19 vaccine hesitancy comments in TikTok videos: A manifest content analysis, Public health nursing (Boston, Mass.). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9874770/.

[2] Ong, J.C. and Cabañes, J.V.A. 2018. Architects of networked disinformation: Behind the scenes of troll accounts and fake news production in the Philippines, ScholarWorks@UMass Amherst. Available at: https://doi.org/10.7275/2cq4-5396.

[3] Chen, S., Xiao, L. and Mao, J. 2021. 'Persuasion strategies of misinformation-containing posts in the social media', *Information Processing & Management*, 58(5), p. 102665. doi:10.1016/j.ipm.2021.102665.

[4] Toquero, C.M. 2022. 'Addressing infodemic through the comprehensive competency framework of Media and Information Literacy', *Journal of Public Health*, 45(1). doi:10.1093/pubmed/fdab412.

[5] Sese, L.V. and Guillermo, Ma.C. 2023. 'E-smoking out the facts: The philippines' vaping dilemma', *Tobacco Use Insights*, 16. doi:10.1177/1179173x231172259.

[6] Cruz, J.C.B., Tan, J.A. and Cheng, C. 2020. *Localization of fake news detection via Multitask Transfer Learning*, *arXiv.org*. Available at: https://doi.org/10.48550/arXiv.1910.09295.

[7] Appendix II: Partial list of web / news / blog sites in the Philippines with fake or unverified content. 2018. Sangguniang Laiko ng Pilipinas. Available at: https://www.cbcplaiko.org/2017/01/31/appendix-ii-partial-list-of-web-news-blog-sites-in-the-philippines-with-fake-or-unverified-content/ (Accessed: 15 December 2023).

[8] Fernandez, A.C. and Devaraj, M. 2019. 'Computing the linguistic-based cues of fake news in the Philippines towards its detection', *Proceedings of the 9th International Conference on Web Intelligence, Mining and Semantics* [Preprint]. doi:10.1145/3326467.3326490.

[9] Posadas-Durán, J.-P. *et al.* 2019. 'Detection of fake news in a new corpus for the Spanish language', *Journal of Intelligent & Fuzzy Systems*, 36(5), pp. 4869–4876. doi:10.3233/jifs-179034.

[10] Bilog, R.J. 2020. 'Application of naïve Bayes algorithm in sentiment analysis of Filipino, English and Taglish Facebook comments', *International Journal of Management and Humanities*, 4(5), pp. 73–77. doi:10.35940/ijmh.e0524.014520.

[11] Fahmida Liza Piya, Rezaul Karim, and Mohammad Shamsul Arefin. 2021. BDFN: A bilingual model to detect online fake news using machine learning technique. *Advances in Intelligent Systems and Computing* (October 2021), 799–816. doi:http://dx.doi.org/10.1007/978-981-16-5301-8_56

[12] Verma, P.K. *et al*. 2021. 'Welfake: Word embedding over linguistic features for fake news detection', *IEEE Transactions on Computational Social Systems*, 8(4), pp. 881–893. doi:10.1109/tcss.2021.3068519.

[13] Waqas Haider Bangyal et al. 2021. Detection of fake news text classification on covid-19 using Deep Learning Approaches. *Computational and Mathematical Methods in Medicine* 2021 (2021), 1–14. DOI:http://dx.doi.org/10.1155/2021/5514220

[14] Z. Khanam, B.N. Alwasel, H. Sirafi, and M. Rashid. 2021. Fake news detection using machine learning approaches. *IOP Conference Series: Materials Science and Engineering* 1099, 1 (2021), 012040. DOI:http://dx.doi.org/10.1088/1757-899x/1099/1/012040