Optimized One Dimensional-Ternary Pattern (1D-TP) for SMS Spam Feature Extraction

Oluwakemi Christiana Abikoye[1], Fatimoh Abidemi Taofeek-Ibrahim[2], Taye Oladele Aro[3]
[1]*Department of Computer Science, University of Ilorin, Ilorin, Nigeria*
[2]*Department of Computer Science, Federal Polytechnic Offa, Kwara, Nigeria*
[3]*Department of Mathematical and Computing Sciences, Kola Daisi University, Ibadan, Oyo State, Nigeria*

Abstract: **Short Message Service (SMS) has become an important form of the mobile communication channel; their popularity is attributed to several conditions including low-cost sending, simple delivery mode and convenient usage. Feature extraction phase has been identified by researchers to be one of the major steps in the Spam SMS detection system. The extraction of features in SMS involves a process of reduction of an initial set of raw features into more manageable forms for processing. This paper employed a new One Dimensional Ternary Pattern (1D-TP) for SMS feature extraction while the simulated annealing was applied to optimize the extracted features. Experimental results showed a better objective function**

Keywords: **Feature Extraction, Short Message Service, Spam SMS, One Dimensional Ternary Pattern,**

I. INTRODUCTION

Short Message Service (SMS) technology evolved out of the GSM Communications standard, an internationally accepted cell phone network specification (Geertsema, Hyman, & van Deventer, 2011). SMS remains the most powerful tool in terms of communication especially for mobile users (Reaves, Blue, Tian, Traynor, & Butler, 2016). It does not limit anyone regardless of high-or low-end mobile phones for as long as they can receive and send messages anytime, anywhere (Shaw & Bosworth, 2012). In a fragmented mobile world of multiple devices, operating systems and service providers, messaging remains the one constant that offers a singular ubiquitous channel through which all end users can communicate with each other (Downer, Meara, Da Costa, & Sethuraman, 2006). SMS grows beyond traditional texting and is now being used in different authentication domains such as mobile banking, one-time password delivery, information retrieval systems, smartphone configuration, Over-The-Air (OTA) configuration and social web site alerts (Mizuki, Matsumoto, Uemura, & Kichimi, 2013).

There is still a major threat of SMS Spam in the mobile communication world despite the different benefits associated with SMS (Abdulhamid, Shafie, Latiff, Chiroma, & Osho, 2017). SMS Spam commonly refers to the unsolicited and unwanted SMS usually conveyed to a large number of recipients (Mahmoud & Mahfouz, 2012). Spam is usually sent in bulk for commercial or other purposes and indiscriminately (Chaudhari, Jayvala, & Vinitashah, 2016) SMS spamming gained popularity over other spamming approaches like email and due to the increasing popularity of SMS communication (Subramaniam, Jalab, & Taqa, 2010). It has become a major nuisance to the mobile subscribers given its pervasive nature. It incurs a substantial cost in terms of lost productivity, network bandwidth usage, management and raid of personal privacy. Mobile Spam frustrates the mobile phone users just like email spam, they cause new societal frictions to mobile handset devices.

The feature extraction phase has been identified by researchers to be one of the major steps in the Spam SMS detection system (Uysal, Gunal, Ergin, & Gunal, 2013). Feature extraction is a process of reduction in which an initial set of raw data is reduced to more manageable groups for processing (Kaur & Rajput, 2013). Dimensionality reduction produces an approximation to original feature in fewer dimensions, while still maintaining the same structure of original features (Telgaonkar, A. H & Deshmukh, 2015). Several feature types in SMS spam filtering have been encountered in different approaches, which have resulted at a different level of accuracies. A content feature usually consists of spam keywords, URL links, monetary value, special characters, emotion symbols and function words. Non-content features consider message metadata such as length, the number of characters, white spaces, the number of terms, date, time and location wise (Zainal, Sulaiman, & Jali, 2015). A vast amount of text classification studies make use of the bag-of-words model to represent text documents where the exact ordering of words, or terms, in the documents is ignored but the number of term occurrences is considered. Each distinct term in a document collection consequently constitutes an individual feature. Terms are assigned particular weights representing their importance in a given document. The most common weighting scheme is Term Frequency - Inverse Document Frequency (TF-IDF) that scales down the number of occurrences of a term in a document by considering the number of documents in the collection containing that term.

The paper introduced a new feature extraction approach for SMS feature extraction, one-dimensional ternary patterns (1D-TP) was used to obtain relevant features from SMS messages. 1D-TP is a statistical technique that was developed on the order of occurrence of the characters (Vikas & Kaur, 2016). The ID-TP patterns were formed from the comparisons of

Unicode values of the characters in SMS messages with the Unicode values of their neighbours. A nature-inspired meta-heuristics optimization algorithm was used to optimize the extracted features from ID-TP. A simulated annealing algorithm was applied to optimize the parameters of ID-TP technique to produce more relevant features.

II. RELATED WORK

Shuaib et al. (2019) proposed the application of a meta-heuristic optimization approach, the whale optimization algorithm (WOA) was used for the selection of salient features in the email corpus and rotation forest algorithm for classifying emails as spam and non-spam. The entire datasets were used, and the evaluation of the rotation forest algorithm was done before and after feature selection with WOA. The experimental results showed that the rotation forest algorithm after feature selection with WOA was able to classify the emails into spam and non-spam with a performance accuracy of 99.9% and a low false-positive rate of 0.0019.

Ma, Zhang, Wang and Chen (2018) presented a new fine categorized SMS spam corpus which they claimed was unique and the largest one to the best of their knowledge. A classifier was proposed based on the probability topic model. The classifier could alleviate feature sparse problem in the task of SMS spam filtering. The system compared the approach with three typical classifiers (k-Nearest Neighbors (K-NN), Naive Bayes (NB) and the Support Vector Machine (SVM)) on the new SMS spam corpus. The experimental results showed that the proposed approach was more effective for the task of SMS spam filtering. However, the Huaiyin Institute of Technology (HIT) SMS Spam Corpus collected was not very enough and has the problem of class-imbalance.

Jantan, Waheed and Ghaleb (2017) applied Feed Forward Neural Network (FFNN) for email spam identification; the weights and biases of this network model were set to optimum using a new modified bat algorithm (EBAT). Experimental results based primarily on two datasets: SPAMBASE and UK-2011 WEBSPAM datasets showed that the developed FFNN model trained by EBAT achieved high generalization performance compared to other optimization techniques.

Adewumi and Akinyelu (2016) integrated firefly algorithm (FFA) with support vector machine (SVM) primarily to develop an improved phishing e-mail classifier (known as FFA_SVM), capable of accurately detecting new phishing patterns as they occur. From a data set consisting of 4,000 phishing and ham e-mails, a set of features, suitable for phishing e-mail detection, was extracted and used to construct the hybrid classifier. Simulation experiments were conducted to evaluate and compare the performance of the classifier. The results produced a classification accuracy of 99.94%, the false-positive rate of 0.06% and a false-negative rate of 0.04%.

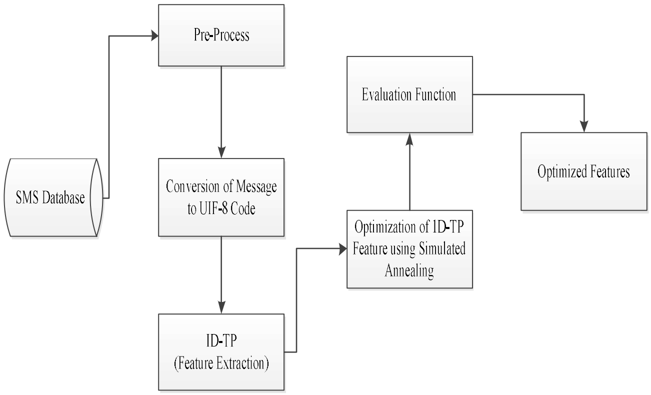
Mustapha and Behjat (2016) developed a new ensemble feature selection techniques for spam detection, based on three feature selection algorithms: Novel Binary Bat Algorithm (NBBA), Binary Quantum Particle Swarm Optimization (BQPSO) Algorithm, and Binary Quantum Gravitational Search Algorithm (BQGSA) along with the Multi-layer Perceptron (MLP) classifier. The experimental results showed accuracy very close to 100% in email spam detection.

Kawade and Oza (2015) used WEKA text classification technique to classify spam message. Different algorithms on SMS dataset were used and based on accuracy, time and error rate a suitable algorithm for the purpose was found. The results of evaluations showed that for different algorithms, accuracy and time are different. It was revealed based on the accuracy and time that Naïve Bayes Multinomial was the best algorithm for classification of spam SMS because its accuracy is highest as well as time required generating model is less than other algorithms.

III. METHODOLOGY

The optimized one-dimensional ternary pattern SMS SPAM detection was developed to improve upon the 1D-TP system of feature extraction approach. Acquisition of datasets was done through the Kaggle Spam SMS publicly available datasets online. The data were preprocessed to remove unwanted characters by stemming and conversion of a message to UTF-8 values of characters in the text using python function. After the preprocessing phase, the preprocessed data was passed into ID-TP feature extraction algorithm. Simulated Annealing was applied to optimize the ID-TP extracted features through parameters setting. Figure 1 shows the framework of the developed system.

Figure 1: Framework of Developed System


Pre-Process
Evaluation Function
Optimized Features
SMS Database
Conversion of Message to UIF-8 Code
Optimization of ID-TP Feature using Simulated Annealing
ID-TP (Feature Extraction)

3.1 Data Preprocessing

The purpose of pre-processing is to transform messages in SMS into a uniform format that can be understood by the learning algorithm. Removing of stop words, words lesser than or equal to two, performing stemming to reduce the vocabulary and converting the remaining part of the message to UTF-8 values of characters in the text, are the basic preprocessing steps to be carried out in this study. Stemming shall be done using the Porter Stemming Algorithm. The algorithm of the Pre-process phase is shown in Figure 2.

```
Step 1: Start
Step 2: Perform stemming
                    /*Porter's algorithm*/
                    from nltk.stem import PorterStemmer
                    from nltk.tokenize import sent_tokenize, word_tokenize

                    ps = PorterStemmer()

Step 3: Convert the message to UTF-8 values of the characters in the text
Step 4: Output the UTF-8 values of the SMS message
Step 5: End
```

Figure 2: Preprocess Phase

3.1 One Dimensional-Ternary Pattern (1D-TP) Algorithm

In 1D-TP, patterns are formed from the comparisons of Unicode values of the characters in SMS messages with the Unicode values of their neighbours. The algorithm of 1D-TP is shown in Figure 3.

```
Step 1: Start
Step 2: Determine the number of neighbours of a character by the parameter P
Step 3:    Assign P/2 characters previously and after of central character, Pc as neighbours of   that character
Step 4:    Compare the value of each member of the 1-D series with its neighbor using the following equation:
        $: \mathrm{TP} = \begin{cases} 1 & \mathrm{Pc} > Pi + \beta \\ 0 & \mathrm{Pc} \leq \mathrm{Pc} + \beta \mathrm{and} \mathrm{Pc} \geq \mathrm{Pi} + \beta \\ -1 & \mathrm{P} < Pi - \beta \end{cases}$
       Where Pc is the centralcharacter,
       β , threshold parameter, which is a user defined parameter
       The value of Pi, localchanges, is within Pc±β
Step 5: The value of the central character (Pc) is adjusted according to ±β value.
         /* Therefore, for each Pc, its neighbors are filtered depending on β*/
Step 6: Two different decimal numbers are generated from the comparison results for each
        Pc
        Negative ones ($\mathrm{Pc} < Pi - \beta$) are employed to generate the low features
        Positive ones ($\mathrm{Pc} < Pi + \beta$)  are used to extract up features
Step 7:  Obtained decimal numbers (up-low) are used instead of the UTF-8 values of
        the text messages
Step 8:  Process continues for each data in the message
Step 9:  Two different 1D signals will be obtained from the up and low values
Step 10:  Histograms which are upper and lower histograms are formed from the 1D signals
Step 11: End
```

Figure 3: 1D-TP Algorithm

3.2 Simulated Annealing Algorithm

The steps involved in the simulated annealing algorithm for optimization is stated in Figure 4.

```
Step 1: Start
Step 2: Set initial value for the maximum possible parameter called the "temperature", as
        Tmax
Step 3: Set initial value for the minimum possible parameter called the "temperature", as
         Tmin
Step 4: Set initial value for the maximum possible Iteration, as MaxIt
Step 5: Set the initial value for the total number of utilized neighbours of characters, i.e. P
        as X1
Step 6: Set the initial value for the threshold parameter, i.e. Beta (B) as X2
Step 7: Start a global Iteration, at T = Tmax  and alpha = random(0, 1)
Step 8: Start a new Local Iteration, at i = 1
Step 9: Compute the cost function (1D-TP transformation of a given SMS message with
         the current values of parameters X1 and X2) as, E = Cost(X1, X2)
Step 10: Generate a random neighboring solution of X1 and X2 as Next_X1 and Next_X2
          respectively
Step 11: Compute the cost function (1D-TP transformation of a given SMS message with
          the next values of parameters Next_X1 and Next_X2) as, E_Next =
           Cost(Next_X1, Next_X2)
Step 12: Evaluate the change in cost as delta_E = E_Next – E
Step 13: if (delta_E  < acceptance_treshold_value) then, move to the next solution by
          accepting the new values of X1 and X2 as, set X1 = Next_X1 and set X2 =
          Next_X2,  goto step 14 otherwise goto step 13 if (Exp(delta_E, T) > random(0,1))
         then, move to the next solution by accepting the new values of X1 and X2 as, set
         X1 = Next_X1 and set X2 = Next_X2
Step 14:  Increment the value of i by 1
Step 15:  Check if the current iteration is the last iteration, i.e. is  i <= MaxIt then goto step
     8       otherwise continue on the next step
Step16: Update the value of T as, set T = alpha * T
Step 17: Check if the current T is out of the region (Tmax, Tmin), i.e. is  T >= Tmin then
           goto step 7 otherwise continue on the next step
Step 18: Output the best solution obtained for X1 and X2 as P and B respectively

Step 19: End
```

Figure 4: Simulated Annealing Algorithm

3.4 Algorithm for the Optimized 1D-TP for SMS Feature Extraction

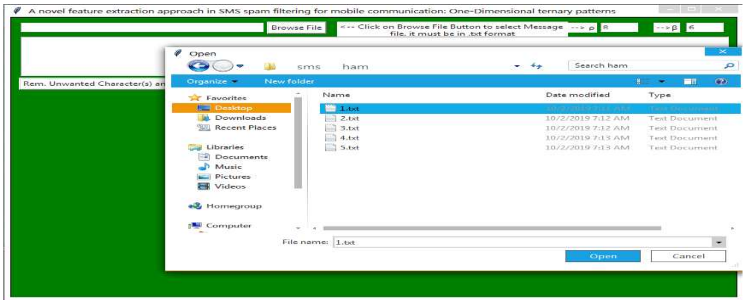
The complete system approach for optimized ID-TP for SMS Spam is shown in Figure 5.

```
Step 1: Start
Step 2: Read in the SMS Message
Step 3: Perform a Preprocessing operation
Step 4: Perform the Feature extraction using 1D-TP (1D-TP transformation&1D-TP
         histogram)
Step 5: Optimization with Simulated Annealing Optimization Algorithm
Step 6: Perform a classification operation
Step 7: End
```

Figure 5: Optimized 1D-TP for SMS-Spam

IV. RESULTS AND DISCUSSION

The sample results of the developed feature extraction approach in SMS Spam filtering for mobile communication using the 1D-TP technique for different HAMs are discussed in these sections. The loading interface is shown in Figure 6.



A novel feature extraction approach in SMS spam filtering for mobile communication: One-Dimensional ternary patterns

<-- Click on Browse File Button to select Message file, it must be in .txt format

--> ρ 8

--> β 6

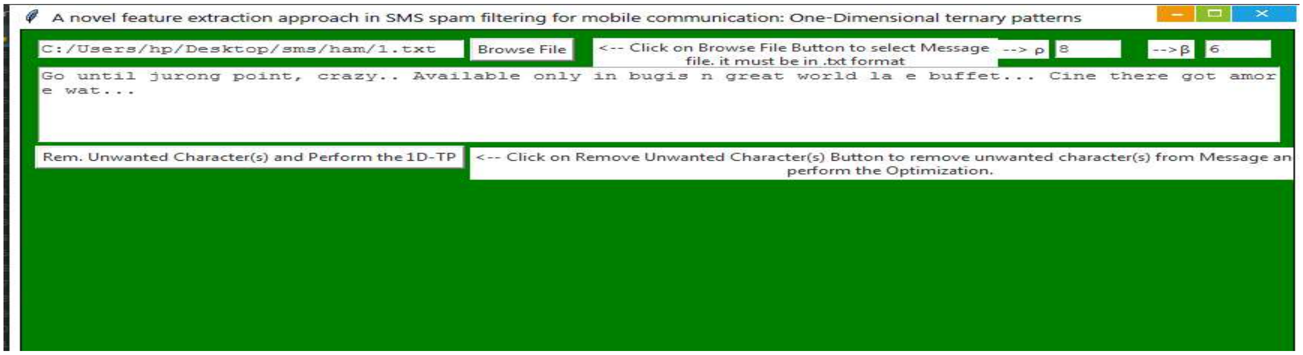
Name	Date modified	Type
1.txt	[illegible]	Text Document
2.txt	10/2/2019 7:12 AM	Text Document
3.txt	10/2/2019 7:12 AM	Text Document
4.txt	10/2/2019 7:13 AM	Text Document
5.txt	10/2/2019 7:13 AM	Text Document

File name: 1.txt

Figure 6: Loading interface of Kaggle SMS Spam data

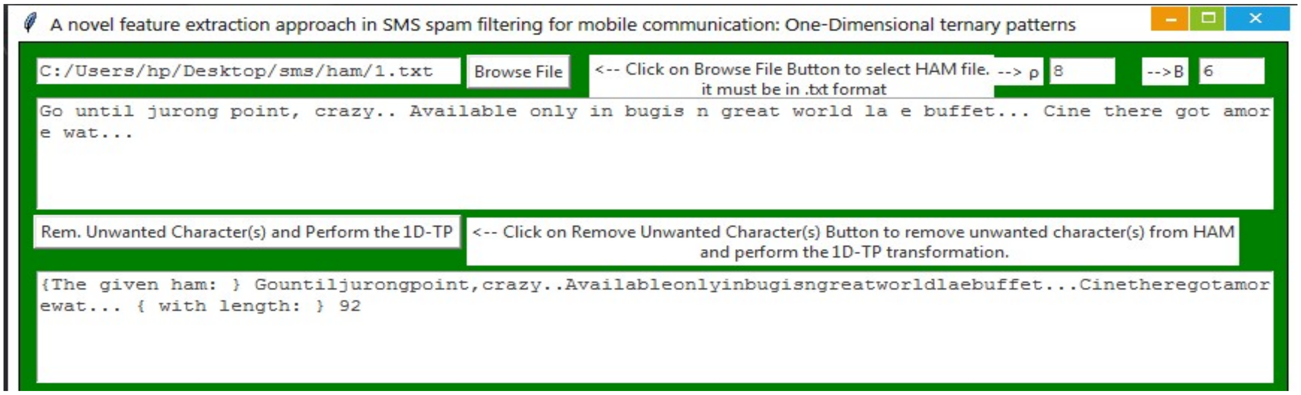
4.1 Results of Preprocess Phase

The wanted characters were removed from the loaded Kaggle SMS Spam dataset. The interface for the removal of unwanted or undesirable characters is shown in Figure 7 and Figure 8.


A novel feature extraction approach in SMS spam filtering for mobile communication: One-Dimensional ternary patterns
Browse File
<-- Click on Browse File Button to select Message file. it must be in .txt format
--> ρ
-->β
Rem. Unwanted Character(s) and Perform the 1D-TP
<-- Click on Remove Unwanted Character(s) Button to remove unwanted character(s) from Message an perform the Optimization.

```
C:/Users/hp/Desktop/sms/ham/1.txt
8
6
Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amor
e wat...
```

Figure 7: Sample of Interface for Pre-processed Kaggle SMS Dataset


A novel feature extraction approach in SMS spam filtering for mobile communication: One-Dimensional ternary patterns
Browse File
<-- Click on Browse File Button to select HAM file. it must be in .txt format
--> ρ
-->B
Rem. Unwanted Character(s) and Perform the 1D-TP
<-- Click on Remove Unwanted Character(s) Button to remove unwanted character(s) from HAM and perform the 1D-TP transformation.

```
C:/Users/hp/Desktop/sms/ham/1.txt
8
6
Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amor
e wat...
{The given ham: } Gountiljurongpoint,crazy..Availableonlyinbugisngreatworldlaebuffet...Cinetheregotamor
ewat... { with length: } 92
```

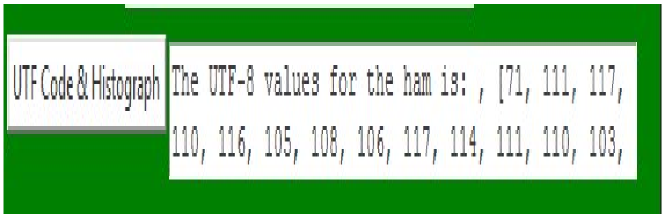
Figure 8 Sample of Pre-processed data

4.2 Results of 1D-TP Feature Extraction (HAM)

The feature extraction outputs are discussed in the following subsections for a sample of HAM

4.2.1 UTF Code and Histogram of HAM

The UTF-8 code and Hstograph for the feature extracted are shown in Figure 9.



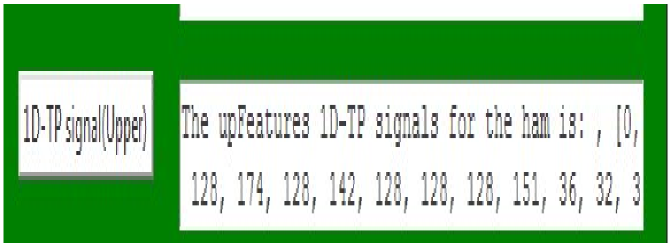
```
The UTF-8 values for the ham is: , [71, 111, 117,
110, 116, 105, 108, 106, 117, 114, 111, 110, 103,
```

Figure 9: Interface of UTF-8 Code (HAM)

From Figure 9, the UTF Code and Histogram produced the lowest value of 71, while 117 was recorded as the highest value of features.

4.2.2 ID-TP Signal Upper Features (HAM)

The values of the result for 1D-TP signal upper features is shown in Figure 11



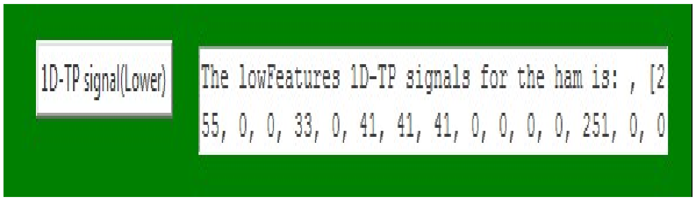
```
The upFeatures 1D-TP signals for the ham is: , [0,
128, 174, 128, 142, 128, 128, 128, 151, 36, 32, 3
```

Figure 10: 1D-TP Signal Upper features

From Figure 10, the 1D-TP signal recorded 0 as the lowest value and the largest number of 174 signal upper features. While 128 has the highest number of occurrence.

4.2.3 ID-TP Signal Lower Features (HAM)

The values of the result for 1D-TP signal lower features is shown in Figure 11.



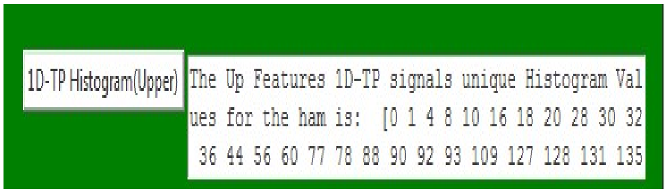
```
The lowFeatures 1D-TP signals for the ham is: , [2
55, 0, 0, 33, 0, 41, 41, 41, 0, 0, 0, 0, 251, 0, 0
```

Figure 11: ID-TP Signal Lower Features (HAM)

From Figure 11, the 1D-TP signal recorded 0 as the lowest value and the largest number of 251 1D-TP signals lower features. While 0 has the highest number of occurrence.

4.2.4 ID-TP Histogram Upper Features (HAM)

The values of the result for 1D-TP Histogram Upper features is shown in Figure 12,



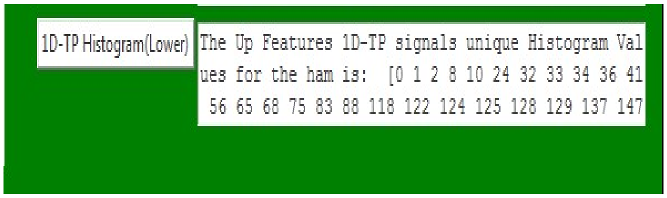
```
The Up Features 1D-TP signals unique Histogram Val
ues for the ham is:  [0 1 4 8 10 16 18 20 28 30 32
 36 44 56 60 77 78 88 90 92 93 109 127 128 131 135
```

Figure 12:1D-TP Histogram Upper features

From Figure 12, the 1D-TP Histogram recorded 0 as the lowest value and the largest number of 251 1DP signal lower features. While 0 has the highest number of occurrence.

4.2.5 ID-TP Histogram Lower Features (HAM)

The values of the result for 1D-TP Histogram lower features is shown in Figure 13.



```
The Up Features 1D-TP signals unique Histogram Val
ues for the ham is:  [0 1 2 8 10 24 32 33 34 36 41
 56 65 68 75 83 88 118 122 124 125 128 129 137 147
```

Figure 13: 1D-TP Histogram Lower Features

From Figure 13, the 1D-TP Histogram Lower features recorded 0 as the lowest value and the highest number of 147 1D-TP Histogram lower features.

4.3 Results of 1D-TP Feature Extraction (SPAM)

The feature extraction outputs are discussed in the following subsections for a sample of SPAM

4.3.1 UTF Code and Histogram (SPAM)

The interface of feature extraction with UTF-8 code and Histograph is shown in Figure 14

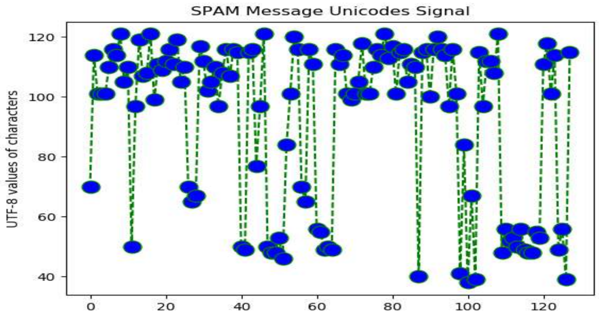


Figure 14: Graphical Interface for UTF Code and Histogram (SPAM)

From Figure 14, the lowest number of UTF-8 values of 70 was recorded, while the highest number of 101 was obtained.

4.3.2 ID-TP Signal Upper Features (SPAM)

The feature extraction interface for 1D-TP signal upper features is shown in Figure 15

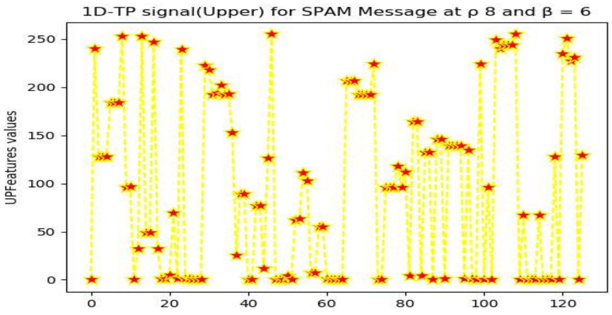
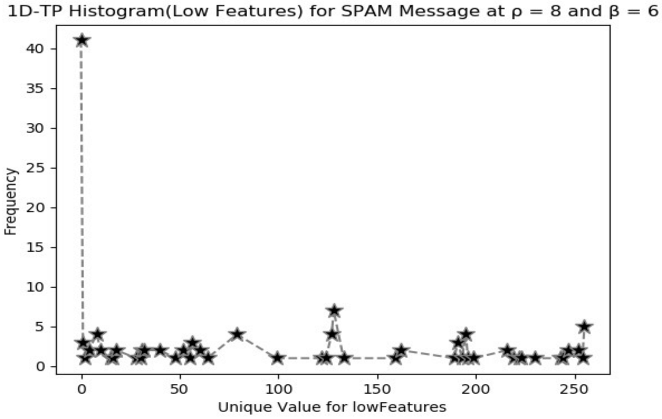


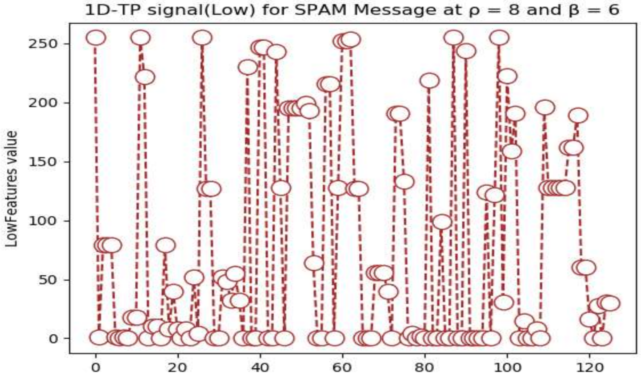
Figure 15: Graphical Interface 1D-TP Signal Upper features (SPAM)

From Figure 15 the 1D-TP signal for upper features recorded 0 as the lowest value and the largest number of 253 upper features. While 128 has the highest number of occurrence.

4.3.3 1D-TP Signal Lower Features (SPAM)



The feature extraction graphical interface for 1D-TP signal lower is shown in Figure 16



.Figure 16: Graphical Interface for 1D-TP Signal Lower Features (SPAM)

From Figure 16, the 1D-TP signal for lower features recorded 1 as the lowest value and the largest number of 255 lower features. While 79 has the highest number of occurrence.

4.3.4 1D-TP Histogram Upper Features (SPAM)

The feature extraction graphical interface for 1D-TP signal lower is shown in Figure 17

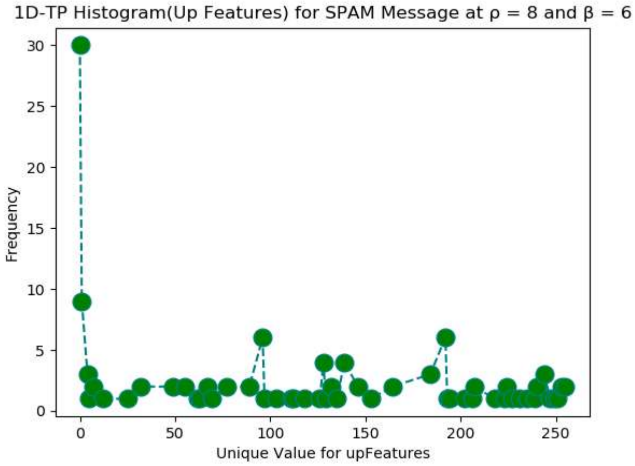


Figure 17: Graphical Interface 1D-TP Histogram Upper features (SPAM)

From Figure 17, the 1D-TP Histogram Upper features recorded 0 as the lowest value and the largest number of 129 lower features.

4.3.5 1D-TP Histogram Lower Features (SPAM)

The feature extraction graphical interface for 1D-TP Histogram Lower is shown in Figure 18

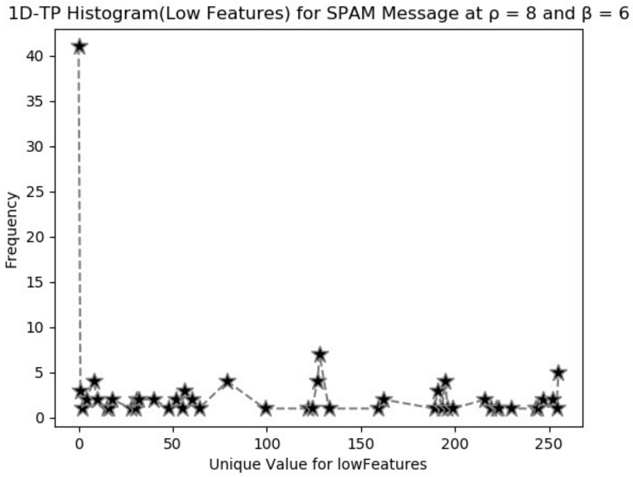


Figure 18: Graphical Interface of 1D-TP Histogram Lower features (SPAM)

From Figure 18, the 1D-TP Histogram Lower features recorded 0 as the lowest value and the largest number of 128 lower features.

4.4 Results of Optimized 1D-TP Feature Extraction algorithm

The section mentions the results obtained for the optimized 1D-TP feature extraction for both HAM and SPAM as shown in subsections 4.4.1 and 4.4.2 respectively.

4.4.1 Optimized 1D-TP Result (HAM)

The best solution with the optimal value of ρ and β was recorded for the best objective 1D-TP of 22. The graphical interface is shown in Figure 19.

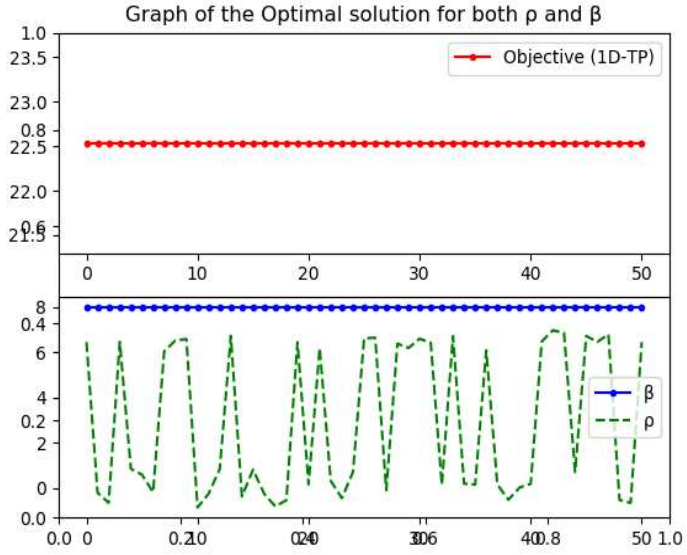


Figure 19: Graphical Interface for Optimal Solution for ρ and β

4.4.2 Result of Optimized 1D-TP (SPAM)

The best solution with the optimal value of ρ and β recorded the best objective function 1D-TP of 24. The graphical analysis is shown in Figure 20.

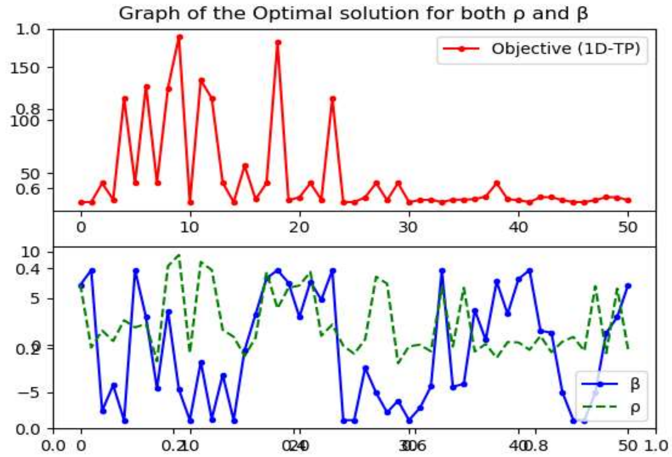


Figure 20: Graphical Interface for Optimal Solution for ρ and β

V. CONCLUSION

Several studies have been conducted in SMS spam detection. Identifying the distinctive features are most likely to be crucial in SMS spam classification. In this paper, a novel approach, 1D-TP is a statistical feature extraction method that is based on the comparisons of characters with their neighbours according to their UTF-8 values was applied to extract features from SMS messages and spam. Kaggle SMS Spam dataset was considered to evaluate the performance of the developed system. Two different features set, low and up features, were extracted from the results of comparisons. The extracted features were further optimized using simulated annealing to select the most relevant and discriminant features.

REFERENCES

[1] Abdulhamid, M., Shafie, M., Latiff, A., Chiroma, H., & Osho, O. (2017). *A Review on Mobile SMS Spam Filtering Techniques A Review on Mobile SMS Spam Filtering Techniques*. (February). https://doi.org/10.1109/ACCESS.2017.2666785

[2] Adewumi, O. A & Akinyelu, A. A. (2016). A Hybrid Firefly and Support Vector Machine Classifier for Phishing Email Detection. *Kybernetes*, *45*(6), 977–994. https://doi.org/10.1108/K-07-2014-0129

[3] Chaudhari, N., Jayvala, P., & Vinitashah, P. (2016). Survey on Spam SMS filtering using Data mining Techniques. *Ijarcce*, *5*(11), 193–195. https://doi.org/10.17148/IJARCCE.2016.51141

[4] Downer, S. R., Meara, J. G., Da Costa, A. C., & Sethuraman, K. (2006). SMS text messaging improves outpatient attendance. *Australian Health Review : A Publication of the Australian Hospital Association*, *30*(3), 389–396. https://doi.org/10.1071/AH060389

[5] Geertsema, S., Hyman, C., & van Deventer, C. (2011). Short message service (SMS) language and written language skills: Educators' perspectives. *South African Journal of Education*, *31*(4), 475–487. https://doi.org/10.15700/saje.v31n4a370

[6] Jantan, A., Waheed, A. H., & Ghaleb, S. A. (2017). Using a Modified Bat Algorithm to Train Neural Networks for Spam Detection. *Journal of Theoretical and Applied Information Technology*, *95*(24), 6788–6799.

[7] Kaur, R., & Rajput, R. (2013). Face recognition and its various techniques : a review. *International Journal of Science, Engineering and Technology Research*, *2*(3), 670–675.

[8] Kawade, D. R., & Oza, K. S. (2015). SMS Spam Classification using WEKA. *International Journal of Electronics Communication and Computer Technology*, *5*, 43–47.

[9] Ma, J., Zhang, Y., Wang, Z., & Chen, B. (2018). A new fine-grain SMS corpus and its corresponding classifier using probabilistic topic model. *KSII Transactions on Internet and Information Systems*, *12*(2), 604–625. https://doi.org/10.3837/tiis.2018.02.004

[10] Mahmoud, T. M. &, & Mahfouz, A. M. (2012). SMS Spam Filtering Technique Based on Artificial Immune System. *International Journal of Computer Science Issues*, *9*(2), 589–597.

[11] Mizuki, A., Matsumoto, T., Uemura, T., & Kichimi, S. (2013). Improving SMS Processing Power for Increasing Smartphone Demand. *NTT DOCOMO Technical Journal*, *14*(4), 60–62.

[12] Mustapha, A., & Behjat, A. R. (2016). Ensemble Feature Subset Selection Technique in Spam Detection System. *ARPN Journal of Engineering and Applied Sciences*, *11*(22), 13135–13140.

[13] Reaves, B., Blue, L., Tian, D., Traynor, P., & Butler, K. R. B. (2016). Detecting SMS spam in the age of legitimate bulk messaging. *WiSec 2016 - Proceedings of the 9th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, 165–170. https://doi.org/10.1145/2939918.2939937

[14] Shaw, R., & Bosworth, H. (2012). Short message service (SMS) text messaging as an intervention medium for weight loss: A literature review. *Health Informatics Journal*, *18*(4), 235–250. https://doi.org/10.1177/1460458212442422

[15] Shuaib, M., Abdulhamid, M., Adebayo, O. S., Osho, O., Idris, I., Alhassan, J. K., & Rana, N. (2019). Whale optimization algorithm - based email spam feature selection method using rotation forest algorithm for classification. *SN Applied Sciences*. https://doi.org/10.1007/s42452-019-0394-7

[16] Subramaniam, T., Jalab, H. A., & Taqa, A. Y. (2010). Overview of textual anti-spam filtering techniques. *International Journal of the Physical Sciences*, *5*(12), 1869–1882.

[17] Telgaonkar, A. H & Deshmukh, S. (2015). Dimensionality Reduction and Classification through PCA and LDA. *International Journal of Computer Applications*, *122*(17), 4–8.

[18] Uysal, A. K., Gunal, S., Ergin, S., & Gunal, E. S. (2013). The impact of feature extraction and selection on SMS spam filtering. *Elektronika Ir Elektrotechnika*, *19*(5), 67–72. https://doi.org/10.5755/j01.eee.19.5.1829

[19] Vikas & Kaur, A. (2016). Face Recognition using Local Ternary Pattern. *International Journal of Science and Research*, *04*(12), 2115–2120.

[20] Zainal, K., Sulaiman, N. F., & Jali, M. Z. (2015). An Analysis of Various Algorithms For Text Spam Classification and Clustering Using RapidMiner and Weka. *International Journal of Computer Science and Information Security (IJCSIS)*, *13*(3), 66–74.