Data Mining Approach to Counter-Terrorism

Uche Stanley Onyekachi[1], Tsopze Norbert[2] & Ebem Deborah Uzoamaka[1*]
[1]Department of Computer Science, University of Nigeria, Nsukka,
[2]Department of Computer Science, University of Yaounde1 , Yaounde, Cameroon
E-mails: onyekachi.uche.pg80062@unn.edu.ng, deborah.ebem@unn.edu.ng, tsopze.norbert@gmail.com
Phones: +2348062256712; +2348052810722; +237 699 517945

ABSTRACT

Terrorism has long been a major threat to the world for many years and different governments have used different approaches to tackle it. This study reports on the use of available data about terrorist incidents all around the world in combating terrorism with the application of deep learning. There was a comprehensive literature review and the analysis of existing systems and ideas gathered were used to develop the system. This project is done in order to improve on the work carried out by Trisha J. (2018). To improve on the work, extra features were introduced in the dataset and a deep neural network (DNN) model for predicting the success of terrorist attacks was developed. Dataset from the Global Terrorism Database (GTD) were used to train the model. Our proposed model achieved performance accuracy of 91.371% as opposed to that of Trisha J. (2018) which achieved the performance accuracy of 91.18%.

Keywords: Terrorism, deep neural network, global terrorism database, data mining

CISDI Journal Reference Format
Uche, S.O., Tsopze, N. & Ebem, D.U. (2020): Data Mining Approach to Counter-Terrorism. Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 11 No 2, Pp 77-90. DOI Affix - https://doi.org/10.22624/AIMS/CISDI/V11N2P5
Available online at www.computing-infosystemsjournal.info

1. BACKGROUND TO THE STUDY

Counter-terrorism is made up of the responses which are taken in order to tackle terrorism [1]. These responses may involve direct military actions in which the military intervenes during terrorist attacks thereby preventing terrorists from attacking [2]. However, the most common approach to counter-terrorism is called “Pre-emptive Neutralization” which mainly involves arresting or even killing of terrorists before they can attack. It also involves questioning of arrested terrorists in order to learn about their future plans, to identify and locate other terrorists. Countries like Israel, United States, United Kingdom, and Russia have been making use of this method of counter-terrorism [3]. Generally speaking, in most of the countries of the world, the task of counter-terrorism is usually the responsibility of trained law enforcement agents [4]. As it stands, due to the recent sophistications in terrorism, there is the need of enhancements in the approaches to counter-terrorism.

Data mining is the process of searching or going through enormous amounts of data in a given set of data, in order to discover information which may be hidden or previously unknown [5]. Through data mining methods, knowledge and patterns could be discovered. The emergences of data mining and automated data analysis methods have become instrumental and resourceful to the law enforcement agents fighting terrorism [6]. The rapid growth of available data in all regions of society requires new computational methods.

Besides traditional statistical techniques [21] and standard database approaches, current research known as Investigative Data Mining (IDM) uses modern methods that originate from researches in Algorithms and Artificial Intelligence. The main goal is the quest for interesting and understandable patterns. This search has always been, and will always be, a critical task in law enforcement, especially for terrorist investigation, and more specific for the fight against terrorism. Examples are the discovery of interesting links between people (social networks [22]) and other entities (means of transport, modus operandi, locations, communication channels like phone numbers, accounts, financial transactions and so on). There are many ways in which IDM can be defined, one of its approach is stated as: "The technique which is used for the organization, sorting, visualization, determining associations and predicting criminal behavior in terrorist networks in order to destabilize them".

IDM differs from traditional data mining applications in significant ways. Traditional data mining is generally applied against large transaction databases in order to classify people according to transaction characteristics and extra pattern in widespread applicability. The problem in IDM is to focus on smaller number of subjects within large background population and identify links and relationships from a far wider variety of activities. In this work deep neural network model was used to predict the success of terrorist attacks. The remainder of this study is organized as follows; section two is literature review, section three is methodology while section four handled the results, section five is discussion and finally section six is recommendation for future work.

2. REVIEW OF LITERATURE

Recently, there have been quite a number of data mining approaches to counter-terrorism. In crime occurrence prediction, [7] was particularly interested in tackling the problems that arose due to high non-linear relationships, redundancies and dependencies between multiple datasets. In order to enhance crime prediction models, [7] considered environmental context information and therefore proposed a feature-level data fusion method with environmental context based on deep neural network (DNN). The dataset used consisted of data obtained from different databases of crime statistics, demographic and meteorological data, and images in Chicago and Illinois. Before generating training data, crime related data were selected through conducted statistical analysis and finally trained the DNN consisting of different layers. Experimental performance results showed that the DNN model is more accurate in predicting crime occurrence than other prediction models.

The limitation of this study is the fact that it is not possible to apply DNN-based crime occurrence prediction method to regions with insufficient data. This lack of data in certain areas may lead to significant performance degradation. In particular, in the absence of crime occurrence report data, crime occurrence prediction is impossible. Again, the crime occurrence prediction method is unable to provide information regarding a specific crime type at a given time slot. In a different study, cross fire attacks were identified by [8] as a recent threat to geographical areas with the objective of disconnecting cities and states from the internet. [8] introduced a framework for timely discovery and detection of attacks during the early stage of the attack known as the "warm-up period of the attack". In order to achieve this, [8] investigated several deep learning methods to uncover the features for attack detection including Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) to detect the attacks during its early period. [8] concluded that "out of the different deep learning methods investigated, the LSTM detection approach achieved encouraging performance under different attack conditions". One limitation of this work is being unable to monitor the performance of the deep learning methods under different structures. A unique deep learning approach to counter-terrorism presented by [9] had a focus on the use of a "recommender system" to detect how terrorists spread online propaganda using different social network media. [9], therefore proposed a "recommender system" for the Global Terrorism Database (GTD) which was based on deep learning.

According to [9], using this recommender system over the GTD will help security agencies for predicting and determining terrorist behaviors. This is achieved by discovering relationships in previous attacks through multi-dimensional visualization of data. An important factor to be considered in applying deep learning models over the GTD datasets is that most of the terrorist attacks occur due to motivations, therefore carrying-out experiments on the GTD datasets is not sufficient. [9], identified the limitation of this work to be the inability to study the behavior of the terrorists over social networks and also added that in order to overcome this limitation, the behaviors of the terrorists should be studied and values are to be assigned for each behavior.

In [10], a project carried out in order to better understand terrorist activities around the world and to use deep learning techniques to predict the success of attacks. [10] used the GTD dataset which consisted of 170,350 terrorist incidents described by a list of 135 features divided into 8 basic groups to predict the success of a terrorist attack using deep learning techniques by exploring different fully connected neural network architectures namely; Fully Connected Neural Network with two hidden layers, Fully Connected Neural Network with three hidden layers and Fully Connected Neural Network with four hidden layers.

The best of the models (Fully Connected Neural Network with three hidden layers) achieved 91.18% accuracy and their results suggested that it is indeed possible to train neural networks to predict the success of a terrorist attack. However [10], identified the limitations as well as avenues for future work on this project viz ; the precision is lower than the recall (from the results of statistical performance on test set) which suggests that the models can still be improved to reduce false positive rate, the use of simple architectures, the use of few/inadequate features in the original dataset, finally there is no “real-world” loss function that takes into account the social, financial and even psychological loss incurred by a false alarm for a terrorist attack. This study is an improvement on [10].

3. METHODOLOGY

Dataset about terrorist attacks from 1970 through 2016 were obtained from the Global Terrorism Database (GTD) maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START). The initial uncleansed dataset consists of 141966 incidents described by a list of 134 features. These features are both categorical and numerical.The dataset is split into input features and the label, the training set, the validation set and the test set. The following tools were used Keras [11-15], TensorFlow [16], Pandas [17], Scikit-learn[18], Matplot lib[19] and Jupyter notebook[20]. We carried three different experiments.

3.1 Exploratory Data Visualization

In order to have a holistic view of the dataset, particularly, the regions under threat of terrorist attacks, it was necessary to see the success of terrorist attacks and the regions involved. From the dataset, the regions that have experienced successful terrorist attacks include: Central Asia, South Asia, Australasia & Oceania, Middle East & North Africa, Sub-Saharan Africa, Eastern Europe, South America, East Asia, Western Europe, Southeast Asia, North America, Central America & Caribbean.

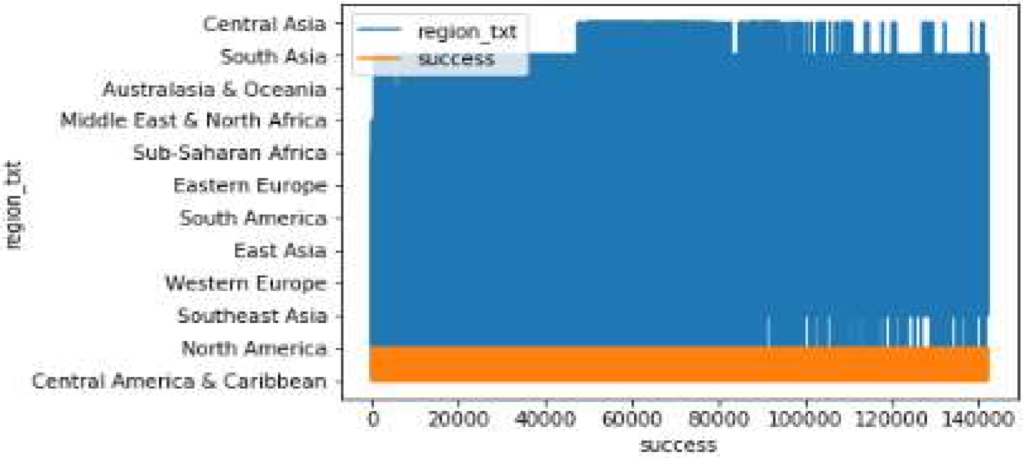


Figure 1 Exploratory Data Visualization

3.2 Data Preprocessing

Data Preprocessing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis. The steps followed to preprocess the dataset include;

3.3.1 Handling the missing data

Missing data can reduce the statistical power of a study and can produce biased estimates, leading to invalid conclusions; therefore, it is important to check for missing values in a dataset. The different ways to handle missing data include; removing the features that have missing data or replacing missing data with another value (usually with zero, the mean, median or the highest frequency value of the given feature). In this work, all the 141966 rows have missing values and 123 features out of the 134 features had missing values greater than or equal to 141000. It was necessary to delete the 123 features with very huge amounts of missing values and 11 features remained without missing values.

3.3.2 Changing categorical data to numerical data

Models can only handle numeric features. At this point, it was very important and necessary to convert the categorical features into numeric features using the “replace” command in python. In this work, the features “country_txt” and “region_txt” had 207 and 12 categorical data respectively, and each of the categorical data was transformed to numeric data by assigning numeric values to them.

3.3.3 Scaling the data

Scaling the data is important in making the input features to have similar orders of magnitude. In this work, the min_max scaler was used to aid the training of our neural network.

3.3.4 Dimensionality reduction

Dimension reduction refers to the process of converting a set of data having vast dimensions into data with lesser dimensions ensuring that it conveys similar information concisely. These techniques are typically used while solving machine learning problems to obtain better features for a classification or regression task.

In this work, the Principal Component Analysis (PCA) was used in dimensionality reduction. PCA is a technique that transforms a dataset of many features into principal components that “summarize” the variance that underlies the data. In PCA, each principal component is calculated by finding the linear combination of features that maximizes variance, while ensuring zero correlation with the previously calculated principal components.

3.3.5 Improvements on the dataset are as follows:

The cleaned datasets possess the following improvements over the initial un-cleaned dataset. There were no incompatible formats of data present in the dataset. The categorical data were transformed to numerical data so that the model could handle them. The dataset preprocessed did not contain missing values therefore there were no errors in the training the model. The dataset features were scaled so that all the dataset will contain features similar in magnitudes, units and range.

3.4 Our proposed deep neural network model

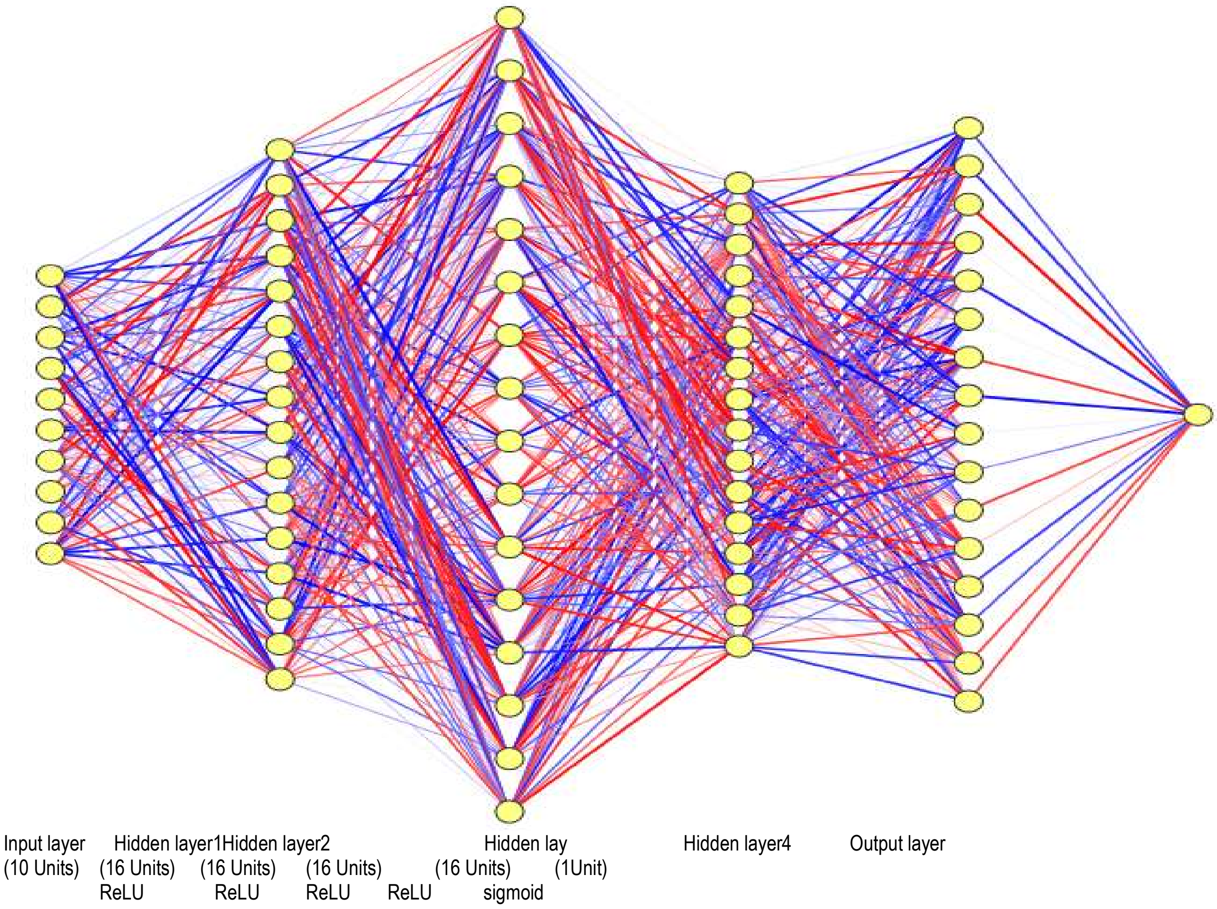


Figure 2: Architecture of the proposed neural network

There are four hidden layers in the proposed model. The input layer has 10 units which correspond to the 10 features of the dataset. The first hidden layer is a dense layer with 16 neurons, ReLU activation function the same with the second, the third and the fourth hidden layers. The output layer is a dense layer with 1 neuron having sigmoid activation function.

4. RESULTS

Performance analysis and experimental results

The metrics defined below will be used throughout this section

i. True positive (TP): Attack data that is correctly classified as an attack
ii. False Positive (FP): Normal data that is incorrectly classified as an attack
iii. True Negative (TN): Normal data that is correctly classified as normal.
iv. False Negative (FN): Attack data that is incorrectly classified as normal.

The following measures were used to evaluate the performance of the proposed model:

a. **Accuracy**
Accuracy is the number of correct predictions divided by the total number of predictions made, given mathematically as (TP + TN) / (TP + TN + FP + FN).

b. **Precision**
Precision is the fraction of relevant instances among the retrieved instances, given mathematically as TP / (TP + FP).

c. **Recall**
Recall is the fraction of relevant instances that have been retrieved over the total amount of relevant instances, given mathematically as TP / (TP + FN).

d. F-score
The F-score measures the harmonic mean of precision and recall which serves as a derived effectiveness measurement, given mathematically as 2 x ((Precision x Recall) / (Precision + Recall)).

Initially, the pandas package (pd) was imported to enable the codes in the package to be used with “pd”. This also made it possible for the dataset to be read and called up into the Jupyter notebook which is the computational environment used in this project. After that, there was the need to use the Scikit-learn package to enable the use of a function called the min-max scaler, which scales the dataset so that all the input features lie between 0 and 1 in order to aid the training of our neural network. In order to build our model, there was also the need to get some codes from the keras package which would enable the full connection of the network. The model was configured using the following settings: optimizer was set to Stochastic Gradient Descent (sgd), Metrics was set to “Accuracy” in order to enable us track the performance accuracy of the model.

To be able to visualize loss and accuracy, there was the need to import the package matplotlib. This package enabled the visualization and graphical representation of the training loss and the validation loss. We carried three experiments to find out the effects of the additional features of the data set and the performance accuracy of our proposed model. The experiments and their results are given below.

First experiment

The first neural network architecture was a fully connected network with 4 hidden layers (figure 2). All the 4 hidden layers consisted of 16 units each with ReLU activation function. The output layer consisted of 1 unit and was set to sigmoid activation function. The model was configured with Stochastic Gradient Descent optimizer and with a batch size of 6400000. After training for 50 epochs, the model achieved 91.371% accuracy on the test set with an accuracy score of 1. This can be visualized graphically in figure 3 below.

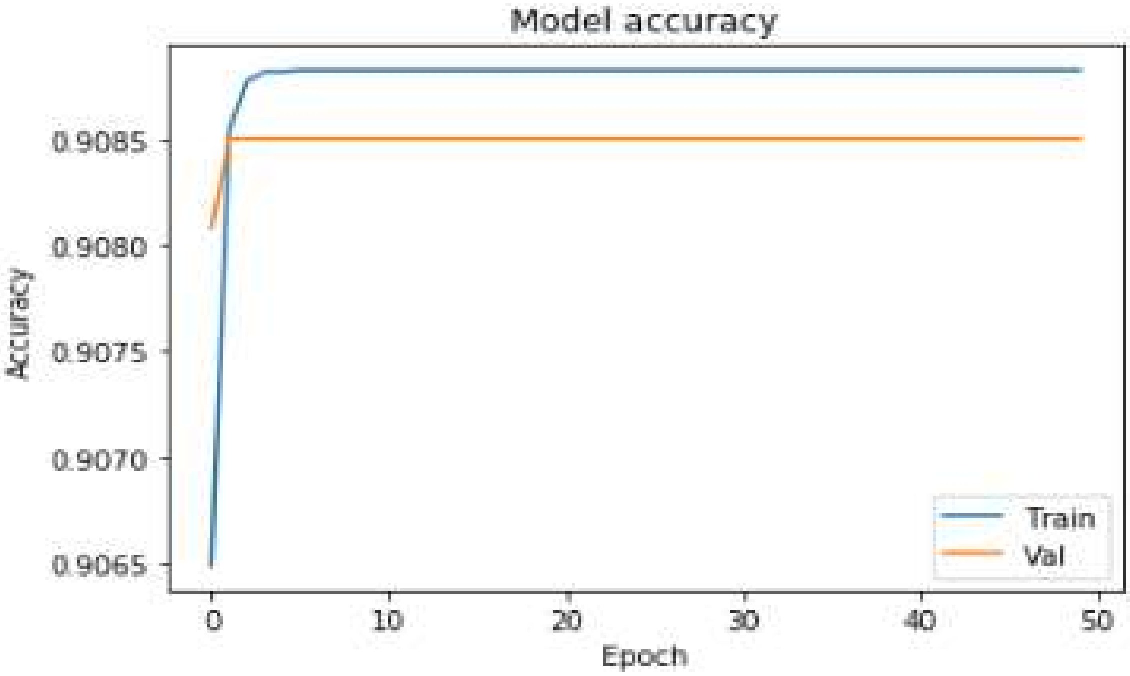
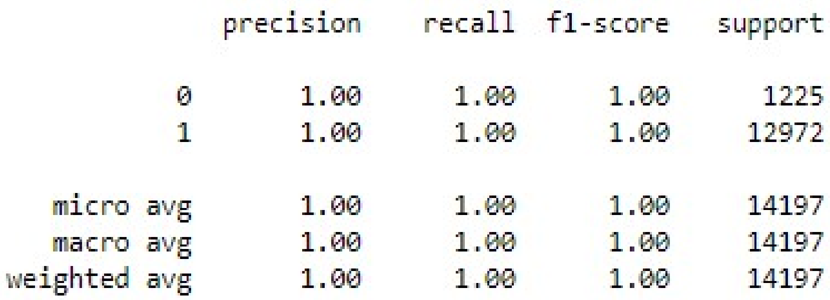


Figure 3: Graphical visualization of proposed model accuracy

The table below shows the classification report on the first experiment.

Table 1: Classification report on the first experiment



```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      1225
           1       1.00      1.00      1.00     12972

   micro avg       1.00      1.00      1.00     14197
   macro avg       1.00      1.00      1.00     14197
weighted avg       1.00      1.00      1.00     14197
```

Second experiment

The second neural network architecture was a fully connected network with 5 hidden layers (figure. 4).

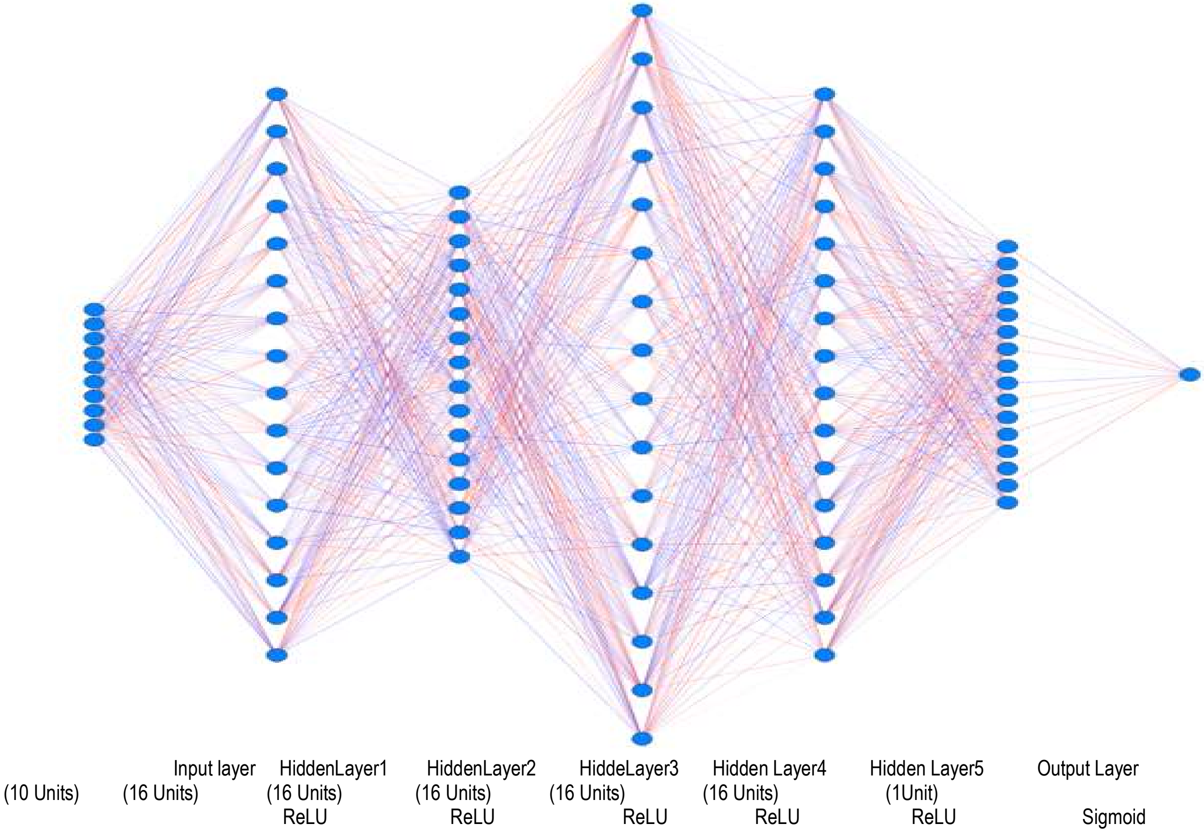

Input layer HiddenLayer1 HiddenLayer2 HiddeLayer3 Hidden Layer4 Hidden Layer5 Output Layer
(10 Units) (16 Units) (16 Units) (16 Units) (16 Units) (16 Units) (1Unit)
ReLU ReLU ReLU ReLU ReLU Sigmoid

Figure 4: Neural network architecture in the second experiment

Batch size was adjusted to 32000. The learning rate was set to 0.0001 and all the hidden layers were set to ReLU activation function and then assigning sigmoid activation for the output layer. After training for 5 epochs, the model achieved 90.935% accuracy on the test set with an accuracy score of 1. This can be visualized graphically in figure 5 below;

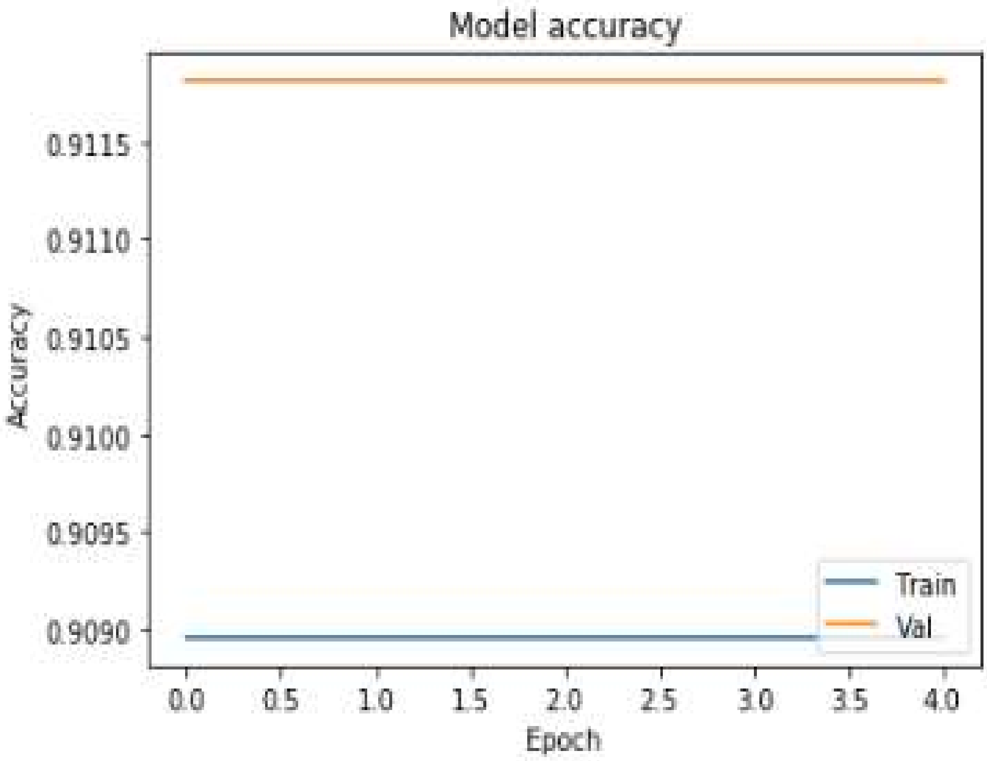
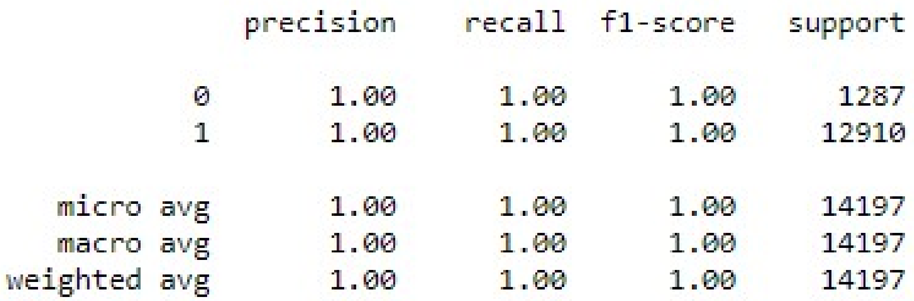


Figure 5: Graphical visualization of second model's accuracy

The following table below shows the classification report on the second experiment.

Table 2: Classification report on the second experiment



```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      1287
           1       1.00      1.00      1.00     12910

   micro avg       1.00      1.00      1.00     14197
   macro avg       1.00      1.00      1.00     14197
weighted avg       1.00      1.00      1.00     14197
```

Third experiment

The third neural network architecture was a fully connected network with 6 hidden layers.

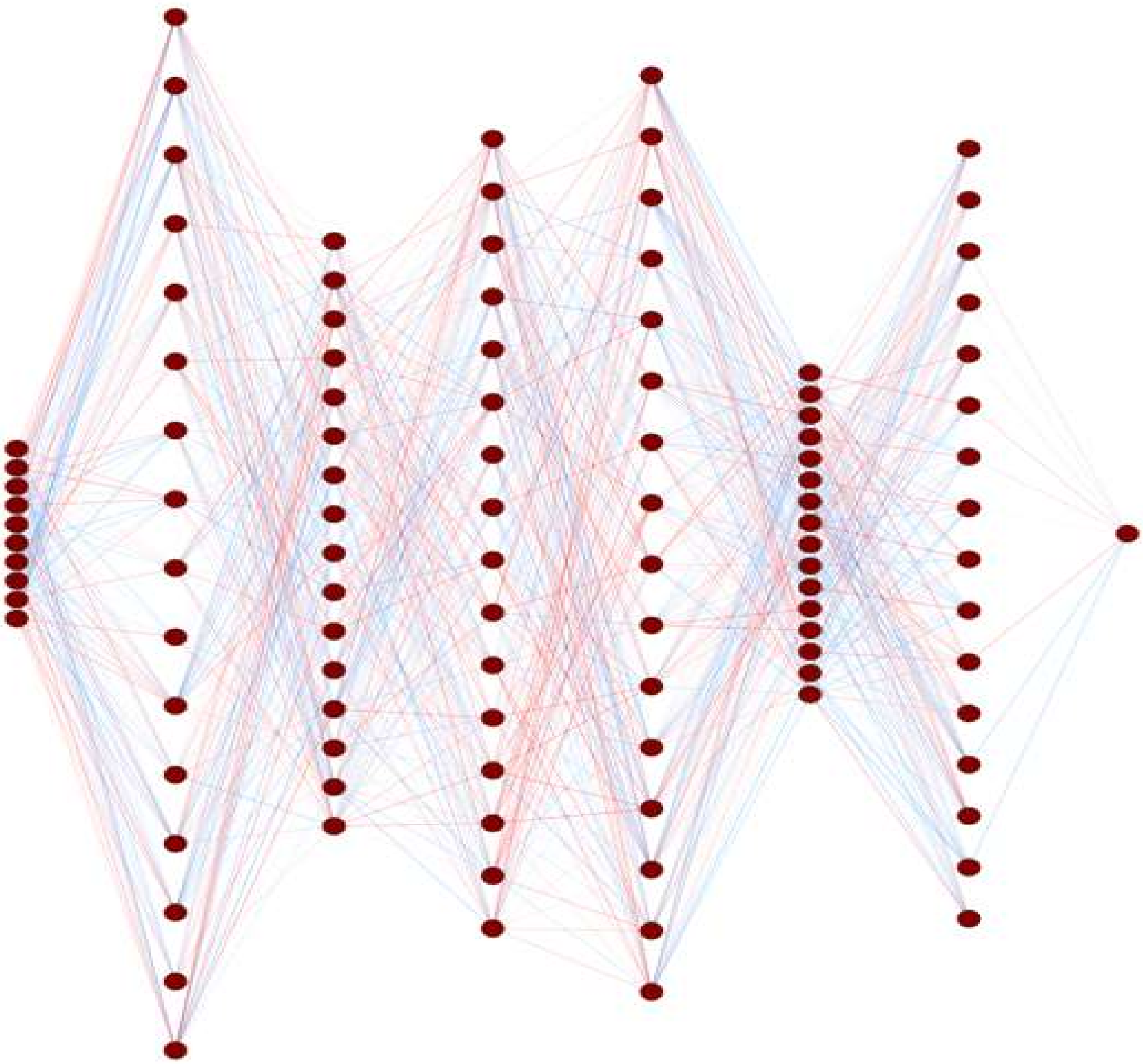

Input layer (10 Units)
HiddenLayer1 (16 Units)
HiddenLayer2 (16 Units) ReLU
HiddeLayer3 (16 Units) ReLU
Hidden Layer4 (16 Units) ReLU
Hidden Layer5 (16 Units) ReLU
Hidden Layer6 (1Unit) ReLU
Output Layer ReLU
Sigmoid

Figure6:Neural network architecture in the third experiment

The model was configured with Stochastic Gradient Descent optimizer and the batch size was adjusted to 32000. The learning rate was set to 0.0001 and all the hidden layers were set to ReLU activation function and then assigning sigmoid activation for the output layer. After training for 5 epochs, the model achieved 90.984% accuracy on the test set with an accuracy score of 1. This can be visualized in figure 7 below.

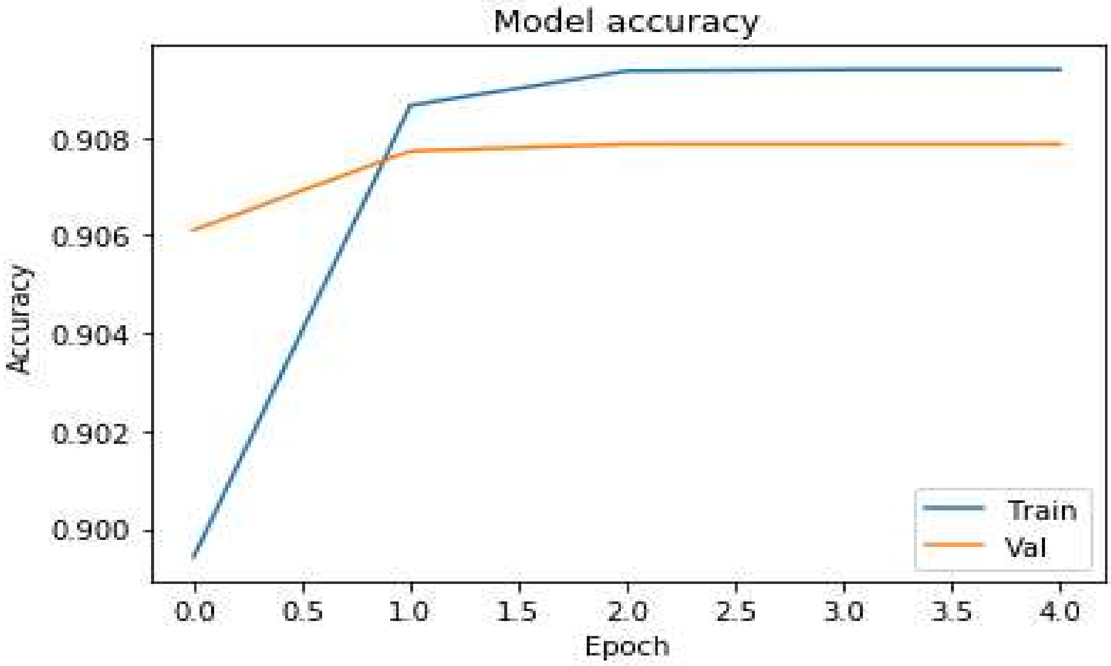
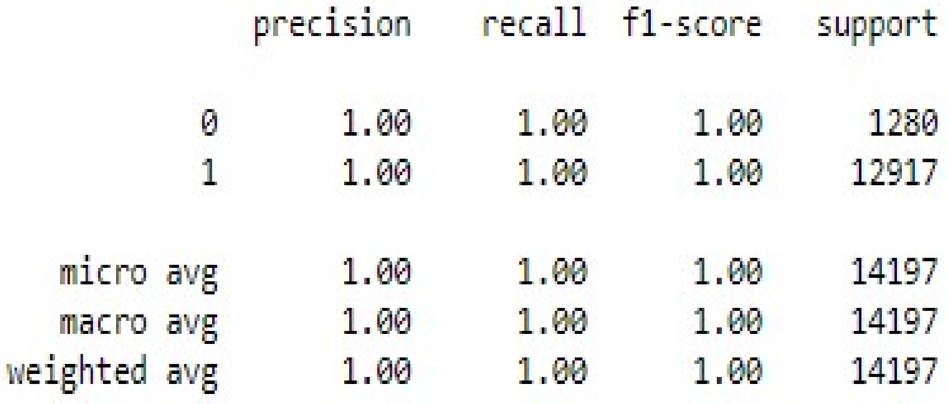


Figure 7: Graphical visualization of third model's accuracy

The following table below shows the classification report on the third experiment

Table 3: Classification report on the third experiment



```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      1280
           1       1.00      1.00      1.00     12917

   micro avg       1.00      1.00      1.00     14197
   macro avg       1.00      1.00      1.00     14197
weighted avg       1.00      1.00      1.00     14197
```

Table 4: Comparison of the Proposed Model to the Baseline Model

Attributes	Proposed model	Existing model
Dataset used	GTD dataset	GTD dataset
Selected features used from this dataset	Event, incident, location, attack, weapon, target, perpetrator, casualties, region, country	Event, incident, location, attack, weapon, target, perpetrator, casualties,
Architecture	DNN (input shape consists of 10 features, four hidden layers of which each of the hidden layers has 16,16,16,16 neurons respectively) and one neuron in output layer.	DNN (input shape consists of 3512 features, three hidden layers of which each of the hidden layers has 100, 50, 20 neurons respectively) and one neuron in output layer.
Activation function	Relu in the hidden layers and sigmoid in the output layer	Relu in the hidden layers and sigmoid in the output layer
Results	Accuracy of 91.371%	Accuracy of 91.18%
	Precision is equal to recall which shows that the model is balanced. Its ability to correctly classify positive examples is the same as its ability to correctly classify negative samples.	Precision is lower than the recall (from the results of statistical performance on test set) which suggests that the models can still be improved to reduce false positive rate.

5. DISCUSSION

Our choice to make use of the proposed model is justified thus; the use of 10 input units which corresponds to the 10 features in the dataset, the use of four hidden layers as opposed to the use of three hidden layers in the baseline model and the use of ten which is an improvement in data preprocessing as opposed to the eight features used in the baseline model.

6. CONCLUSION AND RECOMMENDATION FOR FUTURE WORK

From the results of the experiments on the proposed model, it is possible to improve the performance accuracy of neural network models to predict the success of a terrorist attack. Governments should delve into data mining approaches to counter terrorism because it can provide intelligence to the law enforcement agents as an aid in their counterterrorism measures, it will also reduce the loss of lives of the law enforcement agents responsible for counter-terrorism due to the fact that data mining approach to counter-terrorism reduces the need for physical combat used in preemptive neutralization. Recommendations for future work include a further decrease in the false positive rate by developing a more “real-world” loss function that takes into consideration the social, financial, and even psychological loss incurred by a false alarm for a terrorist attack. This data may be hard to come by but if any estimates exist, it can be used to form a loss function that better mirrors real world circumstances [10].

REFERENCES

[1] Oxford Advanced Learner's Dictionary "International Students Edition" Accessed July 23, 2019.

[2] "Introduction to Foreign Internal Defense" (PDF). Curtis E. Lemay Center for Doctrine Development and Education. Archived from the original (PDF) on January 24, 2017. Retrieved July 10, 2019. [Online].

[3] Savun, Burcu; Tirone, Daniel C. (2018). "Foreign Aid as a Counterterrorism Tool - Burcu Savun,Daniel C. Tirone".Journal of Conflict Resolution. 62 (8):1607–1635. doi:10.1177/0022002717704952.

[4] Stathis N. Kalyvas (2004). "The Paradox of Terrorism in Civil Wars" (PDF).JournalofEthics. 8 (1):97138. doi:10.1023/B:JOET.0000012254.69088.41. Archived from the original (PDF) on October 11, 2006. Retrieved October 1, 2006.

[5] https://www.csis.org/analysis/data-mining-and-data-analysis-counterterrorism . Accessed July 26, 2019.

[6] "Data Mining Curriculum". ACM SIGKDD. 2006-04-30. Retrieved 2014-01-27. Accessed July 26, 2019

[7] H.W.Kang, H.B.Kang, "Prediction of crime occurrence from multimodal data using deep learning" Dept. of Digital Media, Gyonggi-Do, Korea. Accessed on: June 20, 2019. [Online].

[8] M.R.Brust, "Early detection of crossfire attacks using deep learning" Research gate. Accessed on: June 21, 2019. [Online]. Available doi: 10.13140/RG.2.2.23804.6464

[9] R.S.Alhamdani, M.N.Abdullah, I.A.sattar, "Recommender System for Global Terrorist Database Based on Deep Learning" International journal of machine learning and computing, vol. 8, No. 6, December 2018. Accessed on: June 22, 2019.[Online].

[10] Trisha Jani (2018) "Predicting success of global terrorist activities" Accessed on: June 25, 2019.[Online]

[11] "Keras backends". keras.io. Retrieved 2018-02-23. Accessed July 23, 2019. [Online].

[12] "Why use Keras?". keras.io. https://keras.io/why-use-keras/. Accessed July 23, 2019. [Online].

[13] "Core - Keras Documentation". keras.io. Retrieved 2018-11-14. Accessed July 23, 2019.

[14] "Why use Keras?". keras.io. Retrieved 2018-02-23. -11-14. Accessed July 23, 2019.

[15] "Using TPUs | TensorFlow". TensorFlow. Retrieved 2018. Accessed July 23, 2019.

[16] "TensorFlow: Open source machine learning" "It is machine learning software being used for various kinds of perceptual and language understanding tasks" — Jeffrey Dean, minute 0:47 / 2:17 from YouTube clip

[17] "License – Package overview – pandas 0.21.1 documentation". pandas. 12 December 2017. Retrieved 13 December 2017.

[18] Fabian Pedregosa; Gaël Varoquaux; Alexandre Gramfort; Vincent Michel; Bertrand Thirion; Olivier Grisel;Mathieu Blondel; Peter Prettenhofer; Ron Weiss; Vincent Dubourg; Jake Vanderplas; Alexandre Passos; David Cournapeau; Matthieu Perrot; Édouard Duchesnay (2011). "Scikit-learn: Machine Learning in Python". Journal of Machine Learning Research. 12: 2825–2830.

[19] "Matplotlib coding styles". matplotlib.org.Accessed July 23, 2019.

[20] Somers, James. "The Scientific Paper Is Obsolete". The Atlantic. Retrieved 2018-04-10

[21] Heuer, R.: Psychology of intelligence analysis. Center for the study of Intelligence, Central Intelligence Agency (2005)Google Scholar

[22] Pillar, P.R.: Counterterrorism after al Qaeda. The Washington Quarterly 27(3), 101–113 (2004) CrossRefGoogle Scholar