Journal of Engineering Science and Technology
Vol. XX, No. Y (Year) PPP - QQQ

ADOPTION OF ADAM OPTIMIZER FOR ENHANCEMENT OF DEEP LEARNING MODEL IN POLITICAL SECURITY THREAT PREDICTION

LIYANA SAFRA BINTI ZAABAR[1]; SHARIFAH NABILA BINTI S AZLI SHAM[1]; ADRIANA A/P ARUL YACOB[1], A'IN HAZWANI BINTI AHMAD RIZAL[1], KHAIRUL KHALIL ISHAK[2]; MUSLIHAH WOOK[1]; NOR ASIAKIN HASBULLAH[1]; SUZAIMAH RAMLI[1]; MOHD RIZAL BIN MOHD ISA[1]; NOOR AFIZA MAT RAZALI[1,*]

[1]Defence Science and Technology Faculty, National Defence University of Malaysia, Sungai Besi, Kuala Lumpur Malaysia and [2]Center of Cyber Security and Big Data Management and Science University Shah Alam, Selangor, Malaysia
Corresponding Author: noorafiza@upnm.edu.my

Abstract

Political security threats are the main challenges for governments and organizations around the globe and require the development of accurate predictive models for their proactive mitigation. Deep learning techniques have been successful in this area, but optimizing their performance is still a major challenge. Thus, this paper introduces a new way of strengthening deep learning frameworks for the prediction of political security threats by using the Adam optimizer. The Adam optimizer, famous for its efficiency in the optimization of deep neural networks, is used here to improve the predictive capabilities of the existing frameworks. Based on the findings of empirical studies on extensive datasets that cover different political circumstances and types of threats, we show that the proposed approach is effective in increasing the prediction accuracy and model convergence. Besides, the comparative studies with the traditional optimization methods confirm the superiority of the Adam optimizer in the improvement of the performance of deep learning frameworks for political security threat prediction tasks. This research is a step forward in the development of predictive analytics in the political security domains and it shows the importance of the optimization techniques in the improvement of the deep learning models that are used in the real world.

Keywords: ADAM Optimizer; Optimization; Deep Learning; Political Threat; LST

1. Introduction

In recent times, cyberspace has been shown to be greatly capable of affecting national security, and with this observation comes the need to develop more sophisticated defence strategies to address potential risks, as conventional approaches to this issue still struggle to accommodate the vast information exchange facilitated by big data analytics. Within cyberspace, platforms host various diverse data exchanges, including a wide range of public emotional expressions. These emotions can pose security risks, as evidenced by events like the Arab Spring, where negative sentiments were fuelled by misinformation online, which eventually led to societal unrest that threatened national security. As such, promptly detecting disruptive sentiments like these is essential for authorities to effectively manage crises. However, existing methodologies for emotional evaluation regarding national security are inadequate. While most researchers explore various techniques for classifying human emotions, there is insufficient attention given to connecting these emotions to security threats and developing appropriate measurement mechanisms. Despite the capability of sentiment analysis methods to ascertain word polarity, their application in predicting threats remains largely unexplored, particularly in the realm of political security.

Our study aims to remedy this by enhancing the prediction of national security threats, with a specific focus on political security by developing an analytical model by leveraging the Adam optimizer with LSTM deep learning. Our proposed model will contribute to robust capabilities for assessing human emotions and their relationship with security threats. To validate our proposed model, an experimental analysis was conducted. The dataset for this research is constructed by collecting text data from various online platforms, and the proposed model seeks to open new research avenues at the intersection of sentiment analysis and national security. The model will achieve this to enhance emotion measurement and threat prediction in cyberspace.

2. Related Works

This section reviews previous studies on hybrid approaches that integrate lexicon-based methods with machine learning, deep learning, and optimization techniques. These studies provide a foundation for improving the political security model. Sentiment analysis forms the primary basis of this study, as past research has demonstrated that it is a popular technique for identifying threats in the online environment[1].

2.1. Hybrid Lexicon-Based Approach and Machine Learning Technique in Political Domain

Researchers in [2] developed a new theoretical framework to predict political security dangers within cyberspace by combining a lexicon-based approach with machine learning techniques. Within this framework, Decision Tree, Naive Bayes, and Support Vector Machine algorithms are employed as threat classifiers,

and experimental analysis was conducted to validate the framework's efficacy. Their study revealed that the hybrid approach of integrating lexicon-based analysis with the Decision Tree classifier achieves the highest performance in predicting political security threats. These findings contribute valuable insights to the ongoing exploration of opinion mining in threat prediction, and establishes a strong foundation that incentives us to ensure accuracy and diligence in future investigations within the political security field.

2.2. Deep Learning and Optimization Techniques

Deep learning is defined as an influential machine learning (ML) method that studies the properties of different layers or data to provide advanced predictive results [3], and is also a popular method of semi-supervised learning[4]. Optimization is defined as the practice of selecting the most favourable elements from a set of available alternatives. In its basic form, an optimization problem entails maximizing or minimizing a real function by selecting input values from a permissible set, and evaluating the function's value based on those inputs[5].

The author in [6] introduces Gated Attention Recurrent Network (GARN), which combines recurrent neural network (RNN) with attention mechanisms to achieve efficient processing time, complexity, and accuracy in Twitter sentiment analysis. Previously existing literature had failed to achieve this due to large size of Twitter's dataset. GARN resolves this by filtering out low-level and unwanted features, which had not only been reducing the efficiency of classifiers, but was also making up a large bulk of the dataset. This issue of datasets containing large amounts of low-level data is highlighted in [7], where it explains that if this data is left unfiltered, it runs the risk of decreasing the accuracy of sentiment classification, which makes it difficult to obtain an optimal subset of features.

Optimization algorithms, otherwise known as optimizers, are methods used to enhance the performance of deep learning models, as they are able to significantly impact the accuracy and efficiency of a model's training process. During this process, optimizers are responsible for iteratively adjusting the model's weights, while also minimizing the loss function at each epoch.

A comparative study by[8] discusses the various optimization techniques used in deep learning works. ADAGRAD, RMSPROP, ADAM and YOGI make up some of the popular optimizers which are used, as these different optimizers can be utilized to evaluate and experiment with various datasets. The research in [9] introduces two deep neural network models which are based on Convolutional Neural Networks (CNN), and adopts an Adaptive Gradient Algorithm (ADAGRAD) optimizer to address the challenges posed by disaster scenarios. The first model aims to identify tweets which are relevant to disasters, while the second model focuses on extracting detailed information from these tweets, as they can be valuable for various humanitarian activities. These models offer significant contributions to this field of research and the national security domain.

Table 1, which is based on reviewed publications, shows the current adaption of optimization techniques used in various domains. Here, we can see that optimization techniques are experiencing growth. This is due to advancements in algorithms, increased applications, integration with Artificial Intelligence (AI) and Machine Learning (ML), availability of big data and cloud computing resources, and industry demand for improved efficiency. With advancements such

as these, new types of challenges will emerge, and so optimization methods will need to keep evolving in order to mitigate these obstacles.

In [10], researchers proved that the performance of the ADAM optimizer in computational linguistics challenges in neural machine translation from Minangkabau to Indonesian was higher than RMSProp optimizer.

Table 1. Optimization techniques in different domain.

Reference	Technique	Performance Metric
[10]	Computational Linguistics	Adaptive Moment Estimation (ADAM)
[11]	National Security	Adaptive Gradient Algorithm (AdaGrad)
[11]	Online Review/Business	Support Vector Regression With Sequential Minimal Optimization (SMO)
[11]	Computer Sciences	Bitwise Arithmetic Optimization Algorithm (BAOA)

3. Proposed Model

In our research, we developed the ADAM-LSTM Political Security Model to predict various threats from online platforms. Our model comprises of several key stages, including pre-processing, word embedding, threat detection, and parameter tuning. LSTM is used for threat detection, while the parameter tuning is based on the ADAM optimizer. Figure 1 illustrates the Workflow of ADAM-LSTM Political Security Model.

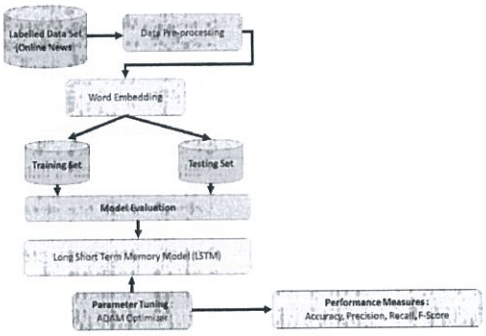

Fig. 1. Workflow of ADAM-LSTM Political Security Model

3.1. Data Collection

In this experimental design, we used the labelled dataset provided by [12]. This dataset was originally created by manually gathering various Malaysian online news sources, such as The Star, New Straits Times (NST), and Free Malaysia Today (FMT), and more. The dataset comprises of 250 texts from online news sources. Out of these texts, 163 of them are categorized as positive, while the remaining 87 are categorized as negative. These positive and negative classifications serve as markers to ascertain the presence of threats within the sample texts.

3.2. Pre-processing

In Natural Language Processing (NLP), text pre-processing is a process that will enhance classifier performance and reduce feature complexity. In this process, unnecessary elements such as punctuation, HTML codes, and symbols are removed, and the gathered text data is then transformed to lowercase and normalized. The normalization process consists of two main steps. First, the unstructured text dataset is converted into a structured word vector, and then, the feature vector's dimensionality is reduced by eliminating unwanted words and stemming them to their original forms. Stemming refers to reducing words to their roots, while lemmatization is the act of utilizing a lexical knowledge base to convert words to their base forms by rooting verbs. At the end of the process, words will be encoded into numerical formats.

3.3. Word Embedding

Word embedding is a technique used in NLP and deep learning to represent words as dense vectors of real numbers[13]. It is a way to map words to vectors in a continuous vector space, where similar words are represented by similar vectors. This experimental layer carries max_words as an input dimension, with 50 as the output dimension (embedding size), and max_len as the input length. This layer creates a low dimensional vector that deals with each word in the input sequences, and directly replaces them with their dense vector representation. These vectors are then multiplied from the embedding layer container, and are then sent to the LSTM layer for further processing.

3.4. Threat Prediction Using LSTM (Long Short-Term Memory)

Authors in[14] developed a global cyber-threat intelligence system using Conventional Neural Network and employed sentiment analysis techniques to detect global threats. In this study, LSTM is used to overcome the problems faced by traditional RNNs in capturing long-term dependencies within sequential data. Traditional RNNS suffer from a vanishing gradient problem, in which gradients will diminish exponentially over time, making it difficult for the network to learn long-range dependencies [15]. LSTM tackles this problem by introducing a memory cell with a more complex structure than the simple recurrent unit used in traditional RNNs[13]. LSTM is driven by the sigmoid neural network layers that regulates the passage of information in and out of the cell[16]. Figure 2 illustrates the ADAM-optimized LSTM Model that combines the ADAM optimization algorithm and LSTM networks for predicting political security threats.

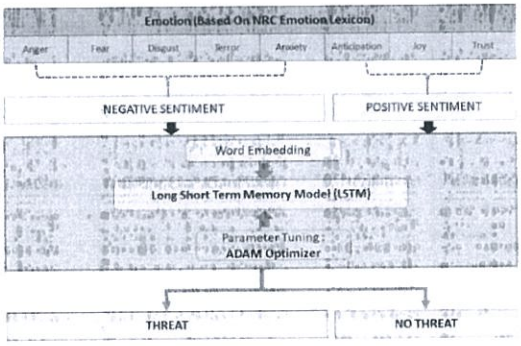

Fig. 2. Political Security Threat Prediction Model

3.5. Parameter Tuning Using ADAM Optimizer

Parameter tuning using the Adam (Adaptive Moment Estimation) optimizer involves adjusting its hyperparameters to improve the performance of a neural network model during training[17]. Adam is a popular optimization method among deep learning model trainers due to its adaptation of gradient descent.

4.Results and Discussion

The technique we developed was tested using the Python 3.11.1 environment, and underwent evaluation on a PC equipped with an Intel® Core™ i7-8650U CPU @ 1.90GHz, 2.11 GHz, 8GB of RAM, a 64-bit OS, and an x64-based processor. We assessed the performance of the ADAM-LSTM approach using a small, labelled dataset of 250 sentences from Malaysian online news sources. To gauge the efficacy of our models, we compared them to the model developed by the researchers cited in reference [2]. We selected their model for comparison because it addresses the same task as our study, which is identifying threats in the political domain, and because it also utilizes the same dataset for evaluation, which is data that is derived from Malaysian online news.

The final phase of this research design is to validate the analysed data. In this phase, this study demonstrated a comparative performance evaluation to validate the proposed theoretical framework. The evaluation test compares the results of precision, recall, accuracy, and F-measure[17].The performance measure involves calculating the accuracy, precision and recall value of the test dataset, and this measure is then used in two phases. The two phases are as follows: Firstly, the evaluations are compared with one another to discern the best optimizer employed in the proposed framework, and secondly, in order to be fully validated, the selected optimizer is compared to the results of isolated deep learning approaches. Evaluation process commenced after the labelling of data into either positive or negative classes, and the imbalance between the class proportions was addressed. A random subset of sentences was selected to train and test the dataset that used the LSTM deep learning technique, followed by the employment of a confusion matrix, which computed the accuracy, precision, and recall of the DL classifiers, allowing for the comparison of algorithmic performance based on training data

labels. Accuracy, defined as the proportion of correctly predicted opinions out of all input opinions to the classifier, is determined by True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values. The formula is as shown in Eq. (1).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision is shown in Eq. (2), and is the percentage of true cases of an opinion (of an instance) among all the classified cases of the opinions (of all instances). To determine the accuracy, true positive rate (TP) was used, as shown in the formula below.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall is defined as the proportion of properly categorised occurrences of a polarity over the total number of correct instances of the polarity. The formula to calculate the recall values using TP and FN is shown in Eq. (3):

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

The F-score is calculated by dividing the number of true positives by the sum of true positives and false positives, as shown in Eq. (4).

$$F - score = \frac{2TP}{2TP+FP+FN} \quad (4)$$

4.1. Comparative output

In Table 2 and Figure 3, the ADAM-LSTM model is compared to other currently existing approaches. The findings indicate that the ADAM-LSTM model surpasses the hybrid methods in performance, and that the single LSTM model yields the least favorable results.

Table 2. Comparative output of the ADAM-LSTM Model with other methods.

Methods	Accuracy	Precision	Recall	F-Score
ADAM-LSTM	94.0%	94.0%	100.0%	96.9%
Untrained LSTM	40.0%	94.7%	38.3%	54.5%
Baseline Model (Lexicon + Decision Tree)	76.0%	92.7%	80.9%	86.4%

Comparing the untrained LSTM to the baseline model reveals distinct performance differences. The untrained LSTM demonstrates notably inferior performance across all metrics in comparison to the trained LSTM, as it lacks any learned knowledge about the data. Meanwhile the baseline model, which employs a Hybrid Approach, exhibits superior performance to the untrained LSTM, although its results still fall short of those achieved by the trained LSTM. The

baseline model has an accuracy of 76 percent, a precision of 92.7 percent, a recall of 80.9 percent, and an F-score of about 86.4 percent, which, while decent, is not the most optimal performance. Lastly, ADAM-LSTM outperforms both untrained LSTM and the baseline model in accuracy, precision, recall and F-score measurements.

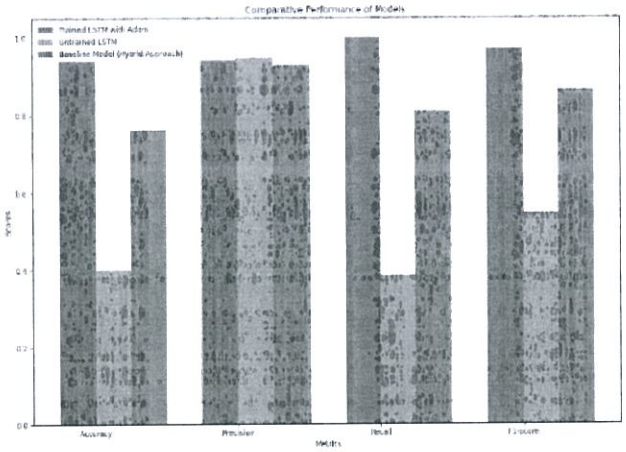

Fig. 3. Comparative Performance of ADAM-LSTM Model With Other Methods

4.2. Area Under Precision Recall (AUC-PR)

In figure 3, an AUC-PR of 0.94 indicates that there is high precision being recalled at different thresholds, suggesting that the model performs well in separating positive classes from negative classes.

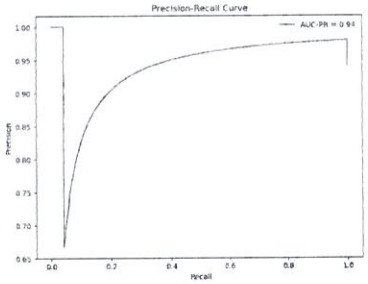

Fig. 5. Precision-Recall Curve of ADAM-LSTM Model

4.3. Training and Validation Accuracy and Loss Curve

The curves in Figure 6 shows the loss and accuracy of the model. Validation Loss and Accuracy are calculated on a separate validation dataset, and serve as indicators of how well the model generalizes unseen data[16]. A decreasing validation loss and increasing validation accuracy throughout the epochs suggest that the model is not over-fitting, and is instead learning meaningful patterns from the data.

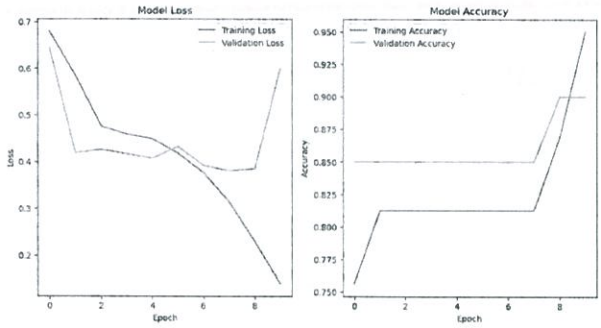

Fig. 6. ADAM-LSTM Model Loss and Accuracy Curves

5. Conclusions

This research study shows that combining the Adam optimizer with LSTM significantly enhances deep learning models' abilities to predict political threats. This model is not only capable of reshaping the political security threat prediction landscape, but can also support researchers' future studies within the national security field. The synergy between the Adam optimizer and LSTM networks offers enhanced accuracy and robustness in national security scenarios by adaptively adjusting learning rates, while also capturing long dependencies that are essential in detecting evolving threats. To summarize, the ADAM-LSTM model introduces new and effective ways to enhance the accuracy, efficiency, and versatility of political threat prediction, making it a significant advancement in the domain of national security.

Abbreviations	
ADAGRAD	Adaptive Gradient
ADAM	Adaptive Moment Estimation
RMSPROP	Root Mean Square Propagation
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory

References

1. M. A. Haq, M. A. R. Khan, & M. Alshehri. (2022). Insider Threat Detection Based on NLP Word Embedding and Machine Learning. *Intelligent Automation and Soft Computing*, 33(1),619–635.
2. Afiza, N., Razali, M., Malizan, N. A., Hasbullah, N. A., Wook, M., Zainuddin, N. M., Khalil Ishak, K., Ramli, S., & Sukardi, S. (2023). Political Security Threat Prediction Framework using Hybrid Lexicon-based Approach and Machine Learning Technique. *IEEE Access*,11,17151–17164.
3. Razali, N. A. M., Malizan, N. A., Hasbullah, N. A., Wook, M., Zainuddin, N. M., Ishak, K. K., Ramli, S., & Sukardi, S. (2021). Opinion mining for national security: techniques, domain applications, challenges and research opportunities. *Journal of Big Data*, 8(1),52.
4. Susmitha, M., & Razia, S. (2023). Utilization of deep learning and semantic analysis for opinion mining in information extraction: a review. *Indonesian Journal of Electrical Engineering and Computer Science*, 30(1), 469.

5. Parmar, M., Agrawal, A., & Tech, M. (2019). A Review On Optimization Techniques Using Data Mining. *International Journal of Research and Analytical Reviews (IJRAR)*, 6(1), 616-621.
6. Parveen, N., Chakrabarti, P., Hung, B. T., & Shaik, A. (2023). Twitter sentiment analysis using hybrid gated attention recurrent network. *Journal of Big Data*, 10(1),50.
7. Ridzwan Yaakub, M., Iqbal Abu Latiffi, M., & Safra Zaabar, L. (2019). A Review on Sentiment Analysis Techniques and Applications. *IOP Conference Series: Materials Science and Engineering*, Joint Conference on Green Engineering Technology & Applied Computing 2019, Bangkok, Thailand, 551(1).
8. Bashetty, S., Raja, K., Adepu, S., & Jain, A. (2022). Optimizers in Deep Learning: A Comparative Study and Analysis. *International Journal for Research in Applied Science and Engineering Technology*, 10(12), 1032–1039.
9. Paul, N. R., Sahoo, D., & Balabantaray, R. C. (2023). Classification of crisis-related data on Twitter using a deep learning-based framework. *Multimedia Tools and Applications*, 82(6), 8921–8941.
10. Almu'iini Ahda, F., Prasetya Wibawa, A., Prasetya, D., & Sulistyo, A. (2024). Comparison of Adam Optimization and RMSprop in Minangkabau-Indonesian Bidirectional Translation with Neural Machine Translation. International Journal on Informatics Vizualization, 8(1), 231–238.
11. Talpur, N., Jadid Abdulkadir, S., Akashah Patah Akhir, E., Hilmi Hasan, M., Alhussian, H., & Hafizul Afifi Abdullah, M. (2023). A novel bitwise arithmetic optimization algorithm for the rule base optimization of deep neuro-fuzzy system. *Journal of King Saud University - Computer and Information Sciences*,35(1),821-842.
12. Ohtomo, K., Harakawa, R., Iisaka, M., & Iwahashi, M. (2024). AM-Bi-LSTM: Adaptive multi-modal Bi-LSTM for sequential recommendation. *IEEE Access*,12,12720–12733.
13. F. Sufi. (2023). A global cyber-threat intelligence system with artificial intelligence and convolutional neural network. *Decision Analytics Journal*, 9(1), 100364.
14. Y. Touzani & K. Douzi. (2021). An LSTM and GRU based trading strategy adapted to the Moroccan market, *Journal of Big Data,* 8(1), 126.
15. Saeed & E. Al Solami.(2023). Fake News Detection Using Machine Learning and Deep Learning Methods, *Computers, Materials and Continua*, 77(2), 2079–2096.
16. Y. Matalon, O. Magdaci, A. Almozlino & D. Yamin.(2021).Using sentiment analysis to predict opinion inversion in Tweets of political communication. *Scientific Report No.* 11.
17. Ali, Md. N. Y., Sarowar, Md. G., Rahman, Md. L., Chaki, J., Dey, N., & Tavares, J. M. R. S. (2019). Adam Deep Learning With SOM for Human Sentiment Classification. *International Journal of Ambient Computing and Intelligence*, *10*(3), 92–116.