

International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 7.056

IJCSMC, Vol. 13, Issue. 10, October 2024, pg.1 – 11

A Secured Keystroke-Based Model for Preventing Social Engineering Attacks using Recurrent Neural Network

Adedeji Adegbenle¹; Oludele Awodele²; Chibueze Ogbonna³; Taiwo Adigun⁴

^{1,2,3}Computer Science Department, School of Computing and Engineering Sciences, Babcock University, Ilishan-Remo, Ogun State Nigeria

⁴Software Engineering Department, School of Computing and Engineering Sciences, Babcock University, Ilishan-Remo, Ogun State Nigeria

¹ adegbenle0099@pg.babcock.edu.ng; ² awodeleo@babcock.edu.ng;
³ ogbonnac@babcock.edu.ng; ⁴ adigunt@babcock.edu.ng

DOI: <https://doi.org/10.47760/ijcsmc.2024.v13i10.001>

Abstract: Among several authentication problems, preventing social engineering attacks using behavioural biometric approach has not received the required attention especially with focus on keystroke dynamics. This study aims to leverage the power of deep learning for more accurate and robust continuous authentication based on typing patterns. The proposed framework for this study utilized deep learning algorithm for behavioural biometrics authentication using Keystroke dynamics. The deep learning model was developed using Recurrent Neural Network (RNN) algorithm and was optimized was to obtain a better performance with Bayesian optimization which, eventually enhanced the model's accuracy. The dataset was split into training and testing in the model design phase and some hyperparameters such as dense, activation, batch size, sigmoid, filament, input size and epoch were used and optimized for building the deep learning algorithm. The RNN model is used to generate the evaluation metrics such as log loss, accuracy, precision and recall. The result presented the accuracy, precision, recall, and loss function as 100%, 100%, 100%, and 36% respectively for optimized model. The cost metrics yielded 0.0032, 0.0032, and 0.0006 MAE, MSE, and RMSE respectively. The developed KBB shows high level of social engineering attacks mitigation in comparison with the existing solution from the performance measure results.

KEYWORDS: Behavioural Biometrics, Keystroke Dynamics, Recurrent Neural Networks, Keylogging, Optimization

I. INTRODUCTION

The current state of traditional authentication methods has made it difficult to ward off the increasing sophistication of cyber-attacks on mobile apps [1]. Aside, from the challenges of preventing online identity theft, cyber fraud, malicious attacks, and phishing in Short Message Service (SMS), and man-in-the-middle attacks, traditional authentication systems are now groping with sterner social engineering attacks. Social engineering attacks are a form of behavioural manipulations which easily bypass access control measures and existing physiological security prevention systems in place today [2]. In social engineering attacks, scammers impersonate trusted officials, like customer service representatives of a bank, to deceive and steal unsuspecting victims of millions of dollars every year [3]. Mobile-based social engineering occurs when attackers make use of real-time payment system opportunities available on the mobile phone to lure users to transfer money or information to them [3]. A vast majority of social attackers make use of email apps on mobile phones and voices to swindle their victims.

Behavioural biometrics identifies patterns in the ways that particular bodies perform particular tasks, that is, patterns in walking, speaking, typing, or even in computer mouse behaviour. These behaviours are as unique as fingerprints and are much harder for malicious actors to capture much less duplicate. These patterns are prohibitively difficult to capture and replicate, and they evolved over time [4]. Unlike traditional authentication methods, which authenticate only when access is initiated, behavioural biometrics technologies authenticate continuously, evaluating a user's ongoing interaction with their computer in real-time [4]. The cumbersome process of traditional biometric authentication requires a good level of attention from the user, like scanning the finger through the fingerprint sensor, placing the camera on the face for facial recognition, and so on. On the other hand, behavioural biometrics authentication, along with geo-location tools uses silent authentication analyses which involve the examination of the behaviour patterns like how they hold their smart devices, how they walk and hold their smartphone, how they swipe their tablet, etc.

The keystroke Behavioural biometrics can leverage powerful statistical models and deep learning algorithms to spot the differences between a known user's gradual evolution and the unwanted presence of an entirely different user [6]. Besides being a potential mechanism for preventing social engineering attacks, it can guarantee a smooth user experience on mobile devices [7]. Behavioural biometrics can achieve a better user experience by collecting large amounts of user data or user parameters from a mobile device and using deep learning to resolve features to match each user.

The proposed framework for this study is to utilize deep learning algorithms for behavioural biometrics authentication using Keystroke dynamics. This study aims to leverage the power of deep learning for more accurate and robust continuous authentication based on typing patterns. Deep learning algorithms, such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs), have shown remarkable success in various tasks, including image recognition, natural language processing, and pattern recognition. By utilizing deep learning models, the study can improve the authentication system's accuracy and adaptability.

II. RELATED RESEARCH

The related works by various researchers on behavioral biometric are elucidated here, and some of the previous studies are used for evaluating the proposed model. [8] developed Active Behavioural Biometric Authentication using CAT swarm optimization variants with deep learning technique. Comparison of several CAT optimization variants for authentication that is active was done, and the fitness functions tested are Rastrigin, Rosenbrock as well as Griewant while the variants of Cat swarm optimization (CSO) considered are AICSO, PCSO as well as ADCSO. An approach to detect social engineering attack in offline texts or real time environments is proposed using CBR and Deep Learning was also developed by [9]. It is a two-stage approach that detects social engineering attacks and is based on natural language processing, case-based reasoning and deep learning. The text is first converted, parsed and checked for grammatical errors using NLP techniques and CBR and then deep learning is used to identify and isolate possible attacks. [10] Considered methods of psychological influence on a person and described the basic techniques used by attackers. The concept of generative-competitive neural networks is considered to evaluate of techniques used by attackers.

Investigation of using CAPTCHA keystroke dynamics to enhance the prevention of phishing attacks is a study by [11]. It is a designed and implemented CAPTCHA keystroke dynamics model in enhancing the prevention of phishing attacks. The evaluation was conducted in a controlled laboratory experiment in order to practically evaluate

the approach applied. [12] developed deep learning keystroke biometrics called TypeNet. It is a novel free-text keystroke biometrics method based on Deep RNN, for authentication and identification at large scale called TypeNet, trained with keystroke sequences from more than 100,000 subjects. The performance is done using three different learning strategies based on traditional classification frameworks (softmax) and Distance-metric approaches (contrastive and triplet loss). A Deep-Learning-Based Approach to Keystroke-Injection was also developed by [13]. It is an Implementation of deep-learning methods such as LSTM and CNN on small datasets to generate payloads for keystroke injection by utilizing a low-computational-power device implanted inside the keyboard. The dataset allowed synthetic keystrokes to be generated directly within a low-computation-power device.

III. BEHAVIOURAL BIOMETRICS

Biometrics is a method of personal identification or authentication that applies pattern recognition techniques to measurable physiological or behavioural characteristics [14]. biometrics deals with measuring parameters connected to life – parameters that are unique to every person and thus identifies said a person with certainty [15]. [16], defined biometric as the study of the automated security technique that uses the human's unique physiological properties like fingerprints, DNA (Deoxyribonucleic Acid), face, hand geometry, and iris, or behavioural properties like a signature, voice texture, and keystrokes, for identification and authentication (verification) purposes. Physiological and behavioural are the two categories of biometric techniques. Whereas the physiological trait seems to be a more stable physical characteristic like hand silhouette, fingerprint, face or back of the eye, and blood vessel pattern in the hand, the behavioural characteristic is a reflection of an individual's mindset [17].

A quantum leap beyond traditional authentications, behavioural biometrics removes any uncertainty about a user's identity whilst preserving privacy – passively and securely [18]. To defeat presentation attacks, a form of a physiological biometric attack where an attacker spoofs or bypassed the authentication with an artificial replica or crafts, e.g., gummy fingers that have fingerprint impressions, or human-based instruments [19], many liveness detection systems were developed. However, they all failed to defend against a particular type of presentation attack called puppet attack, where an attacker places an unwilling victim's finger on the fingerprint sensor [20]. Furthermore, the detection model was trained through Machine Learning methods to enhance the security against fingerprint spoofs [21].

Therefore, as a new generation of user security solutions, behavioural biometrics is used to identify individuals based on the uniqueness of their interaction with electronic devices such as tablets, smartphones, or mouse-screen-and-keyboard [22]. Behavioural biometrics is currently being used in many authentication systems. This is made possible by using machine learning algorithms that keep track of users' unique features and build on such unique features. This involves using user profiles and behaviour such as how the gadgets are helped, how the screen is swiped, and keyboard or gestural shortcuts [23].

A. Types of Behavioural Biometrics

There are several actions that could be taken to uniquely distinguish one person from another. [138], identified three types:

i. Keystroke Biometrics

The behavioural biometric of Keystroke Dynamics uses the manner and rhythm in which an individual types characters on a keyboard or keypad. The keystroke rhythms of a user are measured to develop a unique biometric template of the user's typing pattern for future authentication. Keystrokes are separated into static and dynamic typing, which are used to help distinguish between authorized and unauthorized users. Vibration information may be used to create a pattern for future use in both identification and authentication tasks. Data needed to analyze keystroke dynamics is obtained by keystroke logging [24]. The keyboard is an input device that is used to communicate with the computer by sending input signals to it. While others are fast at typing, some others could be slow, depending on the mood of the user. Hence, biometric keystroke recognition is used to identify a keyboard people use based on how they type rather than what they type.

ii. Signature

Previously, people used their signature as a means of authentication; the way a person's signature appears is dependent on their hand geometry on a writing device. As a behavioural biometric method, a signature could be influenced by a person's emotional or physical condition, and it could be changing periodically. A signature

recognition system can be used to measure the velocity and pressure of the area where the signature was appended on the sensor pad, in addition to taking note of the shape of the signature.

iii. Voice

The mouth, vocal tracts, lips, and nasal cavities are used to create a unique sound, called voice texture in each individual. A person’s voice texture is constant, but its behavioural part may change as time progresses based on the medical condition, age, and emotional stage of the person. Two widely known techniques for categorizing voice recognition are Automatic Speaker Verification (ASV): this uses voice as the authenticating attribute in a two-factor scenario; and Automatic Speaker Identification (ASI). It uses voice to identify who an individual truly is.

B. Advantages of Behavioural Biometric Authentication Systems

There are a few advantages of behavioural biometric systems over traditional manual systems. According to [25], including the advantage of Reduced processing costs and fraud rates,

- i. Reduced error rates, improve convenience,
- ii. Improving scalability,
- iii. Increasing physical safety, and
- iv. Improving accuracy

C. Limitations and Challenges of Behavioural Biometrics Authentication Systems

Despite the many advantages that come with continuous authentication systems, there, however, some challenges that come with it. According to [26], they include Noisy data, Non-universality, Intra-class variations, Lack of uniqueness, Vulnerabilities, Maximizing accuracy, Domain adaptation capability, Biometric feature extraction and selection, lack of real datasets, and Usability. Computation cost and energy consumption.

D. Privacy Concerns of Behavioural Biometric Authentication Systems

Due to security and privacy issues, with people being afraid of breach of trust, behavioural biometrics systems providers seek to address these concerns such as

- i. Users should be educated on the security advantages of this technology using behavioural biometrics to verify their identity.
- ii. Users should be assured of transparency in the usage of their biometric data, and
- iii. Users should have the free will to disembark from any usage of their biometric data.

Therefore, according to [27], when organizations implement behavioural biometrics as a way of validating access control into their space, both physically and on the internet, they need to be cautious of the ethical issues that come with it, because when combined with other data, behavioural biometrics could be used for myriads of disreputable acts due to its covert and frictionless nature.

IV. PROPOSED METHODOLOGY AND EVALUATION

A. Methodology

The framework in Fig.1 explains the principles that guides the procedure of keystroke-based authentication model. The framework also serves as systematic approach to system authentication to ensure accuracy, integrity and consistency of the system. The components include data gathering, data pre-processing, feature extraction, development of neural network model, and Bayesian optimization.

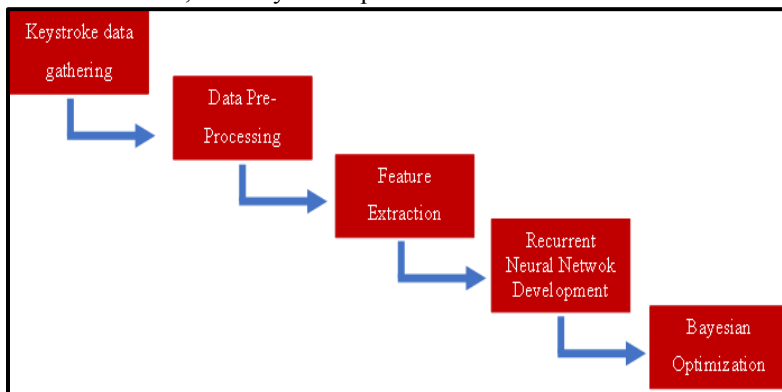


Figure 1: Keystroke Behavioural Biometric (KBB) Development Process

This study creates the Keystroke Behavioural Biometric (KBB) Authentication Model for Preventing Social Engineering Attacks using Bayesian Optimization Technique. This study begins by processing the keystroke dataset, extracting the essence of users' behaviour, building recurrent neural network (RNN) model and obtaining the performance evaluation metrics result with Bayesian deep learning optimization which uncertainty bends to our will, honing the model's accuracy.

i. Dataset Collection

The keylogging dataset was downloaded from UNSW-NB Cyber Range Laboratory of the Australian Centre for Cybersecurity (ACC). The dataset contains 1633 instances. The total number of normal records in the dataset is 1469. The dataset also contains 164 keylogging records and 24 number of columns. Table 2 gives the summary of dataset information.

Table 1: Keylogging Data Description

S/N	Data	Values
1	Number of columns	24
2	Number of normal records	1,469
3	Number of keylogging records	164
4	Total number of records	1,633

ii. Data Pre-processing

The dataset was cleaned to remove irrelevant or noisy data points. Missing value was handled on the dataset by filling the necessary fields using the median of the dataset. The data was scaled, normalized, encoded and missing value handled.

Scaling

Scaling was used to transform the numerical data into specific range so that the algorithm can work effectively with the data. The study utilizes MinMax Scaler approach to scale the data.

$$X_{scaled} = \frac{(X - X_{min})}{(X_{max} - X_{min})}$$

Normalization

Normalization was used to improve the performance and accuracy of the model. It transforms the data features to a common scale in order to reduce the biases in the dataset. Standard scaler was used to normalize the data to avoid overfitting of the model. By taking the mean away and dividing by the standard deviation, it reduces the characteristics.

$$x_n = \frac{x_n - \mu_n}{\sigma_n}$$

Missing Value Handling

The data is checked to ensure that all the values are complete for each instance of the dataset. The data is checked to ensure there are no error in the data entries, empty field. The structure of the dataset is observed to get idea of the values. Isnull() function is used to get the proportion of the missing value. Then, the values of unimportant columns are dropped.

Encoding

This process involves transforming the character field in the dataset to corresponding numerical value so that the deep learning model can carry out effective classification. The protocol names such as TCP, UDP, ARP, IGMP, and RARP in the data set are converted to corresponding numerical values.

iii. Feature Extraction

The next step after the pre-processing is the extraction of visible features from the dataset. Dimensionality reduction approach was used to extract the features from the dataset. This study utilizes the Principal Component Analysis technique to perform dimensionality reduction on the data set. PCA transform high dimensional data into low dimensional data, at the same time preserving the most important features of the data. The formula for PCA is given thus.

$$Cov(X,Y) = \frac{1}{n} \sum_{i=1}^n (x - x')(y - y')$$

iv. Recurrent Neural Network (RNN) Model and Optimization

Building recurrent neural network (RNN) model was to obtain a better performance with Bayesian optimization which, eventually enhanced the model's accuracy. The dataset was split into training and testing in the model design phase and some deep learning hidden layers such as dense, activation, batch size, sigmoid, filament, input size and epoch were used for building the deep learning algorithm, Recurrent Neural Network (RNN). The RNN model is used to generate the performance metrics for evaluation such as log loss, accuracy, precision and recall. The data was split into a training set of 80% and a testing or validating set of 20%.

Bayesian optimization method was adopted for the hyperparameter optimization process in this study due to its ability to modify each epoch's weights during deep learning model optimizers training as well as it brought about minimization of loss function. This algorithm is used to adjust neural network attributes for example; learning rates, number of epochs, number of dense layer units and activation function. Thus, helping in accuracy improvement as well as overall loss reduction. Bayesian optimizer is well appropriate algorithm for classification or regression hyperparameters model optimization. The hyperparameters that were optimized in this study are learning rate, direction, batch size, epoch, and activation.

B. Evaluation Measures

i. Accuracy

One of the performance evaluation metrics in deep learning is accuracy. It is the fraction of predictions our model got right [194]. This is also described as the true negatives and true positives number divided by the true positives (TP), true negatives (TN), False negatives (FN) as well as false positives (FP) number. A TP or TN is an algorithm correctly classified data point while false negative or false positive is an algorithms incorrectly classified data point [195]. The formula for calculating accuracy is given below.

$$\text{Accuracy} = \frac{TN + TP}{TP + TN + FP + FN}$$

ii. Recall

This metrics is also called the true positive rate (TPR). It is the data samples percentage that a Machine Learning or Deep Learning model identifies correctly as belonging to a class of interest which is the positive class out of the whole class samples. The formula for calculating recall is given below.

$$\text{Recall} = \frac{TP}{TP + FN}$$

iii. Precision

This is one of machine learning algorithm's performance indicator, it is the quality of a prediction that is positive made by the algorithm. It refers to the TP number divided by the total positive predictions number that is, TP and FP [196]. It measures a positive classification accuracy of sample. Precision formula is given below.

$$\text{Precision} = \frac{TP}{TP + FP}$$

iv. Mean Absolute Error (MAE)

MAE metric is a very simple one which is the calculation of the absolute difference between predicted as well as actual values. MAE is calculate by summing all the errors and divide them by a total observations number [197]. The formula for calculating MAE is given below.

$$\text{MAE} = \frac{1}{N} \sum |Y - Y'|$$

v. Mean Square Error (MSE)

MSE evaluates the square of the difference between the original values and the predicted values. The gradient values are easily computed in MSE unlike MAE that requires some linear programming tools to accomplish that. The mathematical representation is given below:

$$\text{MSE} = \frac{1}{N} \sum_{j=1}^N (y_j - y'_j)^2$$

vi. Root Mean Squared Error (RMSE)

RMSE also known as root mean square deviation. It is one of the most popularly used metrics for quality of predictions evaluation. It provides how far predictions fall from true values that are measured through the use of Euclidean distance [200]. RMSE can be measured by using the formula below.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N \|y(i) - \hat{y}(i)\|^2}{N}}$$

vii. Logarithmic Loss

Logarithmic Loss is abbreviated as Log Loss. It is also known as cross entropy loss. It is an important metrics in classification problem which are binary. It evaluates the performance of a 5model by weighing the difference between the predicted values and actual values. It works based on the principle that the lower the value of log-loss, the better the performance.

$$\text{Log loss} = -\frac{1}{N} \sum_{i=1}^N (\log(P_i))$$

V. EXPERIMENTAL SETUP, RESULTS AND DISCUSSION

A. Dataset Description

The dataset used on the introduced Bayesian-based RNN model contains 1469 keylogger records, 164 normal records, and 24 columns which were obtained from UNSW-NB database. The keylogging dataset was used to efficiently obtain the proposed model performance. It was split into a training set of 80% and a testing or validating set of 20%. The statistics information of the dataset is described in Table 2.

Table 2: UNSW-NB Dataset Statistical Information

S/N	Statistics	Value
1.	Number of columns	24
2.	Number of rows	1633
3.	Total number of keylogging records	164
4.	Total number of normal records	1469

B. Feature Engineering on the Keylogging Attack Dataset

Feature engineering is done on the dataset to reduce the number of columns in the dataset by choosing the most important features in the dataset using dimensionality reduction technique, normalization is done to normalize the dataset and scale them appropriately.

Dimensionality Reduction

Dimensionality reduction is done on the dataset using principal component analysis (PCA). PCA is used on the dataset to rank the columns present in the dataset based on their important and influence on the deep learning model performance. From the analysis, 15 most important column were selected based on PCA results to carry out the modeling operation. Table 3 shows the PCA results on the all columns.

Table 3: PCA Results on the 23 Columns in the Dataset

S/N	Column Name	PCA Value
1.	Sport	8.984483636
2.	Stime	5.601938036
3.	Bytes	5.500149479
4.	Pkts	4.904953085
5.	Seq	3.42865005
6.	Ltime	1.251295417
7.	Dport	1.021077818
8.	Mean	0.91161932
9.	Dpkts	0.387231068
10.	Sbytes	0.303931881
11.	Max	0.141405754
12.	Dbytes	0.015518912
13.	Srate	2.01409E-15
14.	Drate	-4.23652E-16
15.	Rate	-0.001631127
16.	Spkts	-0.35145568
17.	Sum	-0.760994446
18.	State	-0.851771308
19.	Proto	-1.415025276
20.	Dur	-1.521098736

21.	Flgs	-2.415818592
22.	Stddev	-3.647112133
23.	Min	-3.954203717

From the PCA result shown in Table 3 above, 15 most important columns based on the ranking result is used to carryout deep learning operations in this study. The most important fifteen columns are shown in Table 4.

Table 4: The Most Important 15 Columns

S/N	Columns	PCA Results
1.	Sport	8.984483636
2.	Stime	5.601938036
3.	Bytes	5.500149479
4.	Pkts	4.904953085
5.	Seq	3.42865005
6.	Ltime	1.251295417
7.	Dport	1.021077818
8.	Mean	0.91161932
9.	Dpkts	0.387231068
10.	sbytes	0.303931881
11.	Max	0.141405754
12.	dbytes	0.015518912
13.	Srate	2.01409E-15
14.	drate	-4.23652E-16
15.	Rate	-0.001631127

C. Hyperparameter Tuning Setting

Hyperparameter optimization (Bayesian optimizer) was used to improve the result gotten from parameter tuning optimization. Table 5 illustrates the Bayesian optimizer parameters used and their value.

Table 5: Hyperparameter Tuning Settings

HyperParameter Optimizer	Parameter	Values
Bayesian Optimizer	Number of dense layer unit	1
	Dropout	0.5
	Activation function	Sigmoid
	Learning rate	0.001
	Number of epochs	10
	Input shape	15, 1
	Batch size	32
	Number of trials	15
	Number of iterations	60
	Minimum Hyperparameter units	96
	Maximum Hyperparameter units	480

D. Performance Evaluation Analysis of RNN on Bayesian Optimization

Training of the KBB model with hyperparameter optimization using Bayesian algorithm gives some evaluation results as shown in Table 6. The results suggest an accurate and effective model for preventing social engineering attacks using keystroke dynamics.

Table 6: RNN Metrics Result based on Bayesian Optimizer

S/N	Machine learning Metrics	Result
	Log loss	0.0036
	Accuracy	100
	Precision	1.00

	Recall	1.00
	RMSE	0.0006
	MAE	0.0032
	MSE	0.0032

Figure 2 below give the graphical representation of training loss and validation loss gotten from 10 epoch iterations where line red denote training loss and the blue line represent validation loss and Figure 3 illustrates training and validation accuracy epochs result.

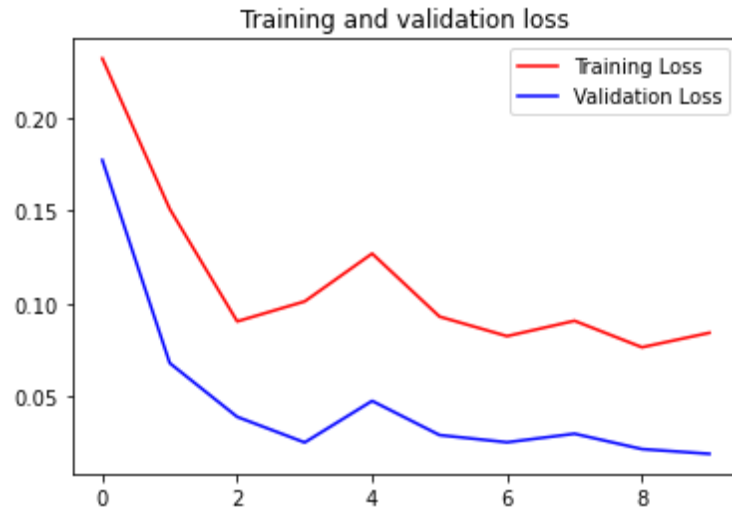


Figure 2: RNN Training and Validation Loss

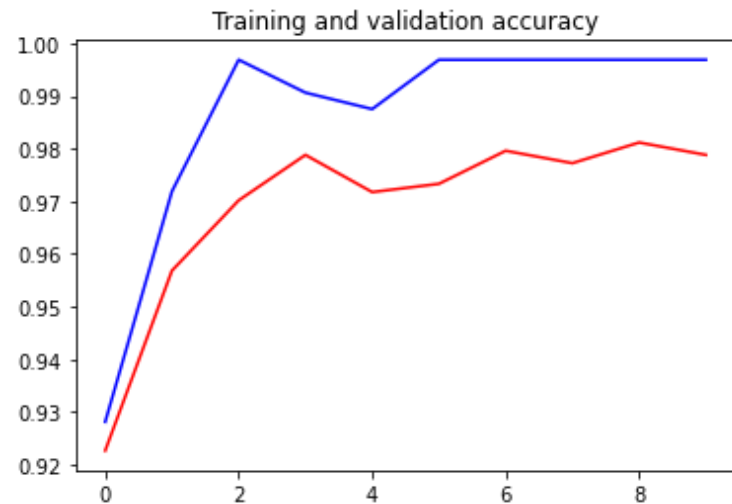


Figure 3: RNN Training and Validation Accuracy

E. Evaluation of Model Performance Against Existing Model

We evaluated the result of KBB in relation to other systems to determine the effectiveness of the solution in the context of algorithm, the modeling approach and social engineering attacks mitigation. Table 7 shows the empirical comparison of evaluation results of the proposed KBB model with other existing solutions. The developed KBB shows high level of social engineering attacks mitigation in comparison with the existing solution from the performance measure results.; where the accuracy result of the proposed KBB model is 99.69% as against the closes t existing solution with 98.29% accuracy.

Table 7: Evaluation of Model Performance Against Existing Model

S/ N	Author	Algorithm	Accuracy Result	Other Metrics Used	Results from other Metrics	SE Attacks Mitigation
1.	[11]	CAPTCHA Keystroke	85	TNR FPR FNR TNR	82.2 0 17.8 100	(CAPTCHAs) keystroke dynamics
2.	[8]	Cat swarm optimization (CSO) considered with RNN	98.29	FAR FRR	0.01 1.02	Hybrid model using typing and mouse information along with facial recognition.
3.	[12]	TypeNet	94.2	ERR Physical ERR Touchscreen	2.2 9.2	TypeNets
4.	[13]	LSTM	96	Precision Recall F1- Score	0.71 0.94 0.83	keystroke-injection payload generation using LSTM Model
5.	KBB Model	RNN with Bayesian	99.69	Log loss Precision Recall RMSE MAE MSE	0.0186 0.9966 1.0000 0.0563 0.0032 0.0108	Keystroke Behavioural Biometric (KBB) Authentication Model.

VI. CONCLUSION

Failure of traditional authentication methods to successfully check the increased trend of social engineering attacks has been portrayed by social attacks or fraud. Measures for developing the keystroke-based biometric authentication model to secure mobile devices and other related systems were provided. Recurrent Neural Network was used for a model to help perform reliable authentication of users of a particular system or mobile devices. This study implements the deep learning model with optimization towards predicting malicious user on a system earlier and preventing them from gaining access to the system. Keystroke-based authentication model was developed to reduce and mitigate the effects of cybercrimes on systems and mobile applications.

Deep Learning models played an important role in mitigating risks of social engineering attacks such as unauthorized access, improper use of users' data etc. Deep learning can be used to develop authorization system which gives no room to cyber attackers, thereby mitigating the social engineering attacks. Particularly, this study has contributed in the determination of best deep learning algorithm to use for authentication model by building a model with high accuracy, therefore proving its reliability compare to previous studies. Also, utilizing Bayesian optimization as the optimization technique on the developed model to enhance performance of recurrent neural network.

REFERENCES

- [1]. Buriro, Z. Akhtar, B. Crispo, and S. Gupta, "Mobile biometrics: Towards a comprehensive evaluation methodology," in Proceedings - International Carnahan Conference on Security Technology, 2017, pp. 1–6, doi: 10.1109/CCST.2017.8167859.
- [2]. L. Razaq, T. Ahmad, S. Ibtasam, U. Ramzan, and S. Mare, "Understanding Mobile-based Frauds Through Victims' Experience," no. April, 2021, doi: 10.1145/3449115.
- [3]. F. Mouton, M. Malan, L. Leenen, and H. Venter, "Social Engineering Attack Framework," no. August, 2014, doi: 10.1109/ISSA.2014.6950510.

- [4]. K. Pfeuffer, M. Geiger, S. Prange, L. Mecke, D. Buschek, and F. Alt, "Behavioural Biometrics in VR Identifying People from Body Motion and Relations in Virtual Reality," Proc. SIGCHI Conf. Hum. Factors Comput. Syst., pp. 1–12, 2019.
- [5]. R. Oak, "A Literature Survey on Authentication Using Behavioural Biometric Techniques," 2018.
- [6]. L. Leonard, Web-Based Behavioral Modeling for Continuous User Authentication (CUA), 1st ed., vol. 105. Elsevier Inc., 2017.
- [7]. E. Ellavarason, R. Guest, and F. Deravi, "Touch-dynamics based Behavioural Biometrics on Mobile Devices – A Review from a Usability and Performance," vol. 53, no. 6, 2020.
- [8]. A. T. Princy and M. K. Preetha, "Active Behavioural Biometric Authentication using CAT Swarm Optimization Variants with Deep Learning," Indian J. Comput. Sci. Eng., vol. 13, no. 3, pp. 653–668, 2022.
- [9]. M. Lansley, N. Polatidis, S. Kapetanakis, K. Amin, G. Samakovitis, and M. Petridis, "Detecting social engineering attacks using case-based reasoning and deep learning," CEUR Workshop Proc., vol. 2567, pp. 39–48, 2019.
- [10]. N. Ryabchuk et al., "Artificial intelligence technologies using in social engineering attacks," CEUR Workshop Proc., vol. 2654, pp. 546–555, 2020.
- [11]. E. K. Alamri, A. M. Alnajim, and S. A. Alsuhibany, "Investigation of Using CAPTCHA Keystroke Dynamics to Enhance the Prevention of Phishing Attacks," Futur. Internet, vol. 14, no. 3, pp. 1–21, 2022, doi: 10.3390/fi14030082.
- [12]. A. Acien, A. Morales, J. V. Monaco, R. Vera-Rodriguez, and J. Fierrez, "TypeNet: Deep Learning Keystroke Biometrics," IEEE Trans. Biometrics, Behav. Identity Sci., vol. 4, no. 1, pp. 57–70, 2022, doi: 10.1109/TBIOM.2021.3112540.
- [13]. V. Gurcinas, J. Dautartas, J. Janulevicius, N. Goranin, and A. Cenys, "A Deep-Learning-Based Approach to Keystroke-Injection," Electronics, vol. 14, no. 13, pp. 1–29, 2023.
- [14]. J. Wayman, A. Jain, D. Maltoni, and D. Maio, "An Introduction to Biometric Authentication Systems," Biometric Syst., pp. 1–20, 2005, doi: 10.1007/1-84628-064-8_1.
- [15]. C. Otti, "The past, present and future of biometrics," Int. J. Eng., pp. 163–168, 2017.
- [16]. S. Shrivastava, "Biometric : Types and its Applications," no. April, pp. 10–11, 2015.
- [17]. H. Gamboa and A. Fred, "A Behavioural Biometric System Based on Human Computer Interaction," Biometric Technol. Hum. Identif., vol. 5404, pp. 381–392, 2004, doi: 10.1117/12.542625.
- [18]. J. Junquera-s, C. Pages-ar, and C. Cilleruelo, "Comparing Machine Learning Classifiers for Continuous Authentication on Mobile Devices by Keystroke Dynamics," 2021.
- [19]. I. 30107-1, "ISO_IEC 30107-1_2016(en), Information technology — Biometric presentation attack detection — Part 1_ Framework." 2016.
- [20]. C. Wang, Y. Wang, Y. Chen, H. Liu, and J. Liu, "User authentication on mobile devices: Approaches, threats and trends," Comput. Networks, vol. 170, pp. 1–26, 2020, doi: 10.1016/j.comnet.2020.107118.
- [21]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.
- [22]. A. Verma, V. Moghaddam, and A. Anwar, "Data-Driven Behavioural Biometrics for Continuous and Adaptive User Verification Using Smartphone and Smartwatch," Sustain. Eng. Sci., vol. 14, no. 12, pp. 1–17, 2022.
- [23]. T. Eude and C. Chang, "One-class SVM for biometric authentication by," no. February, pp. 1–16, 2017, doi: 10.1111/coin.12122.
- [24]. K. Corpus, R. Joseph, D. Gonzales, L. Veja, and A. Morada, "Mobile User Identification through Authentication using Keystroke Dynamics and Accelerometer Biometrics," pp. 11–12, 2016.
- [25]. J. Pato, L. Millett, and W. Biometrics, Biometric recognition, vol. 4, no. 8. 2010.
- [26]. A. Serwadda, V. Phoha, Z. Wang, R. Kumar, and D. Shukla, "Toward robotic robbery on the touch screen," ACM Trans. Inf. Syst. Secur., vol. 18, no. 4, 2016, doi: 10.1145/2898353.
- [27]. I. Traore, I. Woungang, M. S. Obaidat, Y. Nakkabi, and I. Lai, "Online risk-based authentication using behavioral biometrics," Multimed. Tools Appl., vol. 71, no. 2, pp. 575–605, 2014, doi: 10.1007/s11042-013-1518-5.