International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

BERT Model Based Identification and Classification of Web Vulnerabilities Using Deep Learning Approach

Mr. Manjunatha K M*[1], Dr. M Kempanna[2], Mrs. Pushpa G[3], Mr.Rangaswamy M G[4].

Submitted: 03/02/2024 **Revised**: 11/03/2024 **Accepted**: 17/03/2024

Abstract: In recent years, researchers have been focused upon machine learning and machine language based models to predict and identify effects of their researches. In this research the vulnerabilities in web, using the machine learning model BERT (Bidirectional Encoder Representations from Transformers) with additional layers have been attempted. The datasets used for the model's prediction and classification are *SQLInjection* (SQLI) (namely: attacks and benign) and Cross Site Scripting (XSS) datasets respectively. The developed BERT model predicts the vulnerabilities in the data and classifies them accordingly. The loss is estimated through cross entropy loss technique. The performance of the model is evaluated through metric evaluation method namely binary accuracy. The analyses and findings shows that the developed advanced BERT obtained higher accuracy (SQLI with 98% and XSS with 97% accuracies respectively), than the standard BERT model (SQLI with 87% and XSS with 84% accuracies respectively). The research concludes stating that an increased BERT layers based model performs significantly with higher accuracy in classification than the standard BERT as a transformer model.

Keywords*: Bidirectional encoder representations from transformers, BERT, SQL Injections, SQLIA, cross site scripting, XSS, transformers.*

1. Introduction

The web attacks like broken authentication, cross site scripting, SQLInjections and more have been in existence over decades (Bisht et al., 2010; Balasundaram and Ramaraj, 2011; Rahman et al., 2017). Though it is considered as high risk in the usages of technology implemented areas like hospitality sector, educational sector, health sector, IT sector, financial sector and more, they have been causing negative impact upon both government (public) and private businesses entities (Ross, 2018). As the technology advances the attacks has been modified, developed, adjusted and evolved too (Fang et al., 2018). The current lack of defence system as issue has urged the algorithm developers and researchers to mitigate strategies and improvise solutions towards existing web vulnerabilities for the companies to handle the issues and attacks (Gong et al., 2019; Abdulmalik, 2021). Devoting time, budget, human resources and money to secure official data, personal data and other documents in the web has been considerably increased as the attacks increased (ADC, 2015).

Among the common web attacks, there are few that are recurring and commonly identified by the application defence centre and cyber crime units, namely cross site scripting (XSS), SQLInjections attacks (SQLIA), sensitive data exposures, broken authentication, security misconfiguration, insecure direct object references, cross-site request forgery (CSRF), missing function-level access-control, known vulnerabilities based component attacks, un-validated forwards and redirects and more (Chen et al., 2020).To prevent these attacks and damages and secure data from third parties and hackers, companies have turned towards website attack identification and detection models (prediction models) as strategic mechanism (Azman et al., 2021). The prediction models in machine learning are a statistical technique to analyse and predict future forecasts through data mining (Kumar, 2011; Kumar and Binu, 2018).The machine learning models are categorized into three types supervised (needs labeled data for training), unsupervised (determines hidden and underlying patterns) and reinforcement (reward and punishment: action system).There are different prediction models like decision trees, neural networks and regressions that are adopted based on the data usage, research purpose and relevant studies/models (Barde, 2020).

Each machine learning (ML) model is developed based on the necessity of the researcher's purpose or action (Alghawazi et al., 2022). Currently "Transformers" a deep-learning based architecture that has been recently gaining more popularity among the prediction and classification

[1] Senior Scale Lecturer,Department of CSE,Government Polytechnic, Tumakuru,Karnataka,India
[2] Associate Professor, Department of AI &ML,Bangalore Institute of Technology, Bengaluru,Karnataka,India(Affiliated to VTU,Belgavi)
[3] Senior Scale Lecturer,Department of CSE,Government Polytechnic, Channasandra,Bengaluru,Karnataka,India
[4] Senior Scale Lecturer,Department of CSE,Government Polytechnic, Turuvekere,Karnataka,India
E-mail Id: [1] manjunathkm009@gmail.com, [2] kempsindia@gmail.com ,[3] pushpakanth25@gmail.com,[4] mail2mgr.swamy@gmail.com
** Corresponding Author: Mr.Manjunatha K M*
Email: manjunathkm009@gmail.com

ML applications (Ross, 2018). The Transformers are the neural network based that learns and examines the contexts by breaking series of words as relationships into data sequences (Merritt, 2022). Transformers are self-attention and significantly weighing the different input parts that also includes recursive outputs. The Transformer models use the computer vision (CV) and natural language processing (NLP) (Lee et al., 2021). It includes both encoder and decoder where based on the research purpose the model could be adjusted. Among the Transformer models in the machine leaning, BERT (Bidirectional Encoder Representations from Transformers) has been gaining more focus since it is a significant breakthrough in ML, especially the NLP application (Lee et al., 2021).

The BERT model is made of a transformer with encoder for its architecture and doesn't have any decoder. The BERT is developed by Google-AI. The BERT mainly uses NLP (Natural Language Processing) application. The NLP task namely language translation, sentiment analysis and more uses the BERT model. Since the model is monitored and controlled by itself, it falls under the 'self-supervised' machine learning category of algorithms in machine language. BERT's core is based on Transformers model and thus it tokenizes the string values (words) into sub-words and numerals that are processed and served as input values in the predictive models. To predict the website vulnerabilities, researchers use different models like neural network (NN), decision trees (DT), logistic regressions (LR), gradient boosted (GDB), random forest (RF), transformers and more (Farooq, 2021). The current research uses the deep learning based transformer model to study and examine the textual datasets towards identifying and predicting website vulnerabilities.

1.1 Research objectives

The study aims at detecting the web vulnerability (i.e. attacks) through developing a detection and prediction model. The objectives of the study are:

1. To develop a machine learning model to detect and classify the SQLInjection (SQLI) as web vulnerability;
2. To develop a machine learning model to detect and classify the Cross Site Scripting (XSS) as web vulnerability;
3. To identify the best model between the base BERT and advanced BERT by evaluating performances (through metric evaluation) and retaining the model with the highest accuracy achieved.

Through these objectives the research intends to develop, identify, predict and compare the outcome by processing the datasets with base BERT model and the advanced Bert model.

1.2Research purpose

The current research analyses and examines the vulnerabilities in the web using the textual datasets and thus BERT is found significant. In this research the deep learning based transformer is used since it is found to be more rapid, robust, reliable and significant in examining contextual datasets. In 2018, Google AI language developed the BERT model to study textual datasets which is found to be more reliable and robust than other ML models that are presumably effective in non-contextual datasets. By using two different datasets the research develops an advanced BERT model in identifying and classifying the textual datasets.

2. Literature Review

The surplus information through internet as resource has been attacked and hacked by the third parties at large to gain personal information of business clients in private and government websites. The language to build a website can be penetrated and attacked once the vulnerabilities are found by the attacker/hacker. The top web attacks are SQLIA and XSS. The study by 'PreciseSecurity.com' revealed that cyber attacks in the year 2019 used XSS majorly (40%) for attacking the websites by injecting malicious scripts into the most trusted websites (Chen et al., 2020). Similarly in 2022, a global study conducted by 'Statista' found that 33% attacks are made by SQLIA and 26.7% by the XSS, respectively. A study by 'InvictiSecurity.com' in 2021 revealed that the SQLIA was found to be at large where, 32% victims are the government organizations and 35% are the educational institutions and the rest 33% includes private health sector, financial institutions, hospitality and tourism sector, IT sector. It could be understood that educational sector and government organizations are not worried about the data security and web vulnerabilities like other sector and fail to invest towards data security.

Authors Bogale and Tamiru examined and studied about the web vulnerabilities using the datasets SQLIA and XSS. They developed a ML model using the secure hash algorithm-512. They found that NLP is effective and significant in identifying the textual contents in ML model. Similarly, they used a reinforcement model instead of supervised or unsupervised in which the model self-learned and prevented false alarms. Study by Sukhanand and Sharma (2017) used evolutionary fuzzing interface model by using the SQLI and XSS datasets. The study found and concluded that web attacks and web vulnerabilities can be detected precisely by the static examination, programming security analysis and hybrid ML models.

Table 1: SQLI and XSS datasets in prediction of web vulnerabilities

S. No	Author	Year	Model	Datasets
1	Bogale and Tamiru	2021	Secure Hash Algorithm (SHA512) based design science-research method (DSRM) approach	XSS and SQLI attacks
2	Sukhanand and Sharma	2017	Evolutionary Fuzzing based interface model	SQLI and XSS
3	Johari and Sharma	2012	AES (Advanced Encryption Standard) and RSA (Rivest-Shamir-Adleman) algorithms based CNN (Convolution NN) model	XSS and SQLIA
4	Ross	2018	J48, Support Vector Machine (SVM), Artificial NN (ANN), Random Forest (RF) and JRip algorithms based models	Web-app, Correlated and Datiphy based SQLI datasets
5	Azman et al.,	2021	Web-app (Knowledge-based and Signature-based) model	SQLIA datasets

Johari and Sharma in 2012 analysed and examined about convolutional NN (CNN) based architecture in identifying and classifying multiple datasets using hybrid model. The study's findings stated that by improvising the layers in the adopted model, the performance and accuracy can be increased too. Similarly, to classify multiple datasets a model with advanced algorithm is found significant and effective. Author Ross in 2018 used four different algorithms in his model and classified the web attacks using the SQLI datasets. Though different algorithms were used the models were separate and not a hybrid model; henceforth the study concluded the best algorithm among the adopted techniques. The author concluded stating that, by using different architecture than CNN and multiple classifications, the textual classification can achieve greater accuracy. Author Azman (2021) used SQLIA datasets using a web-app model.

Though all these models used SQLI and XSS datasets, it was found through the review that either a hybrid model or an advanced model is preferred for contextual text classification in the ML.

Thus from the literature reviews under focus, it is observed and deduced that, developing a contextual text classification model in ML is effective in identifying and classifying the web vulnerabilities. Among the other ML models BERT model is found more reliable and significant. Similarly, among web attacks and web vulnerabilities the SQLIA and XSS are found to be the most common attacks. Thus in this research, SQLI and XSS are used for identifying and classifying the web vulnerabilities using the BERT model. However, the BERT developed here will be adjusted and improvised with more additional layers to improve performance and increase the accuracy of classification. Similarly to prevent from overfitting/under-fitting issues using the dropout layers in the ML models is proved effective (Srivatsava et al., 2014).

3. Proposed Model

The BERT (Bidirectional Encoder Representations from Transformers) is a language model originally developed by the Google-AI. Though the BERT is considered as a neural-network (NN) based architecture it is popularly known among the researchers as Recurrent NN (RNN) architecture. Mostly many transformer models functions as an encoder-decoder model. In this research, the model unlike other models functions as encoder since the research uses the BERT architecture. BERT encodes the information and doesn't decode the information. Once the encoding is done, the BERT generates a model for data processing. The BERT includes two major steps namely:

- '*pre-training'* in which the unlabeled datasets of the models are trained with different tasks and
- '*fine-tuning'* where the pre-trained parameters of the model are initialized using downstream based labelled datasets.

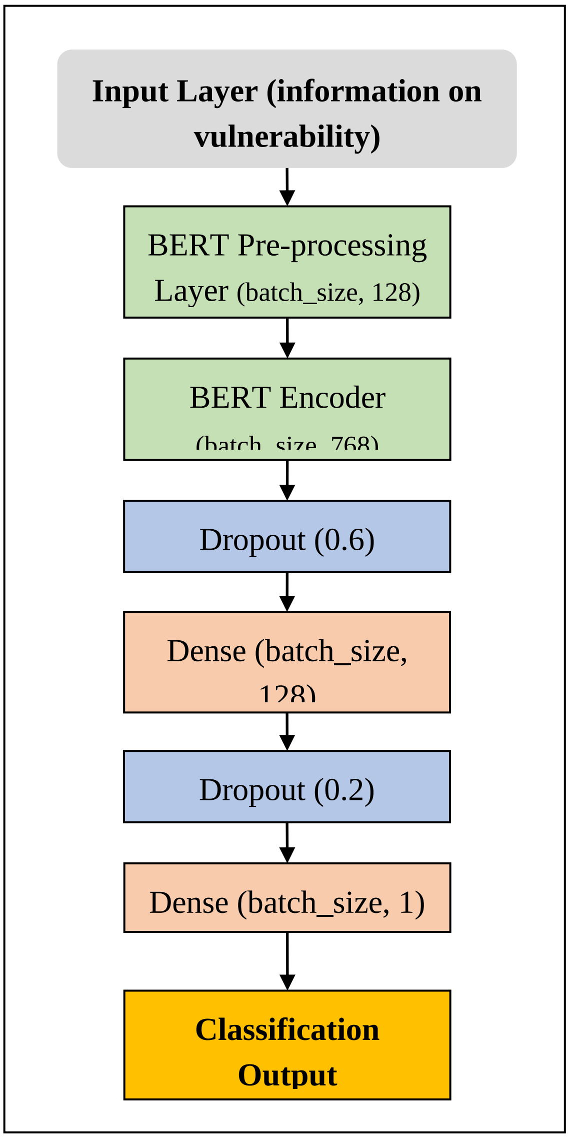
Input Layer (information on vulnerability)
BERT Pre-processing Layer (batch_size, 128)
BERT Encoder (batch_size, 768)
Dropout (0.6)
Dense (batch_size, 128)
Dropout (0.2)
Dense (batch_size, 1)
Classification Output

Fig 1: BERT text-model

The base of the BERT (BERT $_{\text{Base}}$) includes twelve layers of transformer-blocks containing 768 as its hidden size; twelve attention-heads and 110million parameters, which are trainable. Followed by the BERT Large includes, 24-layers of transformer-blocks with sixteen attention-heads and lastly 340 million parameters that are trainable (refer to figure 1). Basically it could be viewed as a three layer architecture with input and output as separate layers (refer to figure 2).

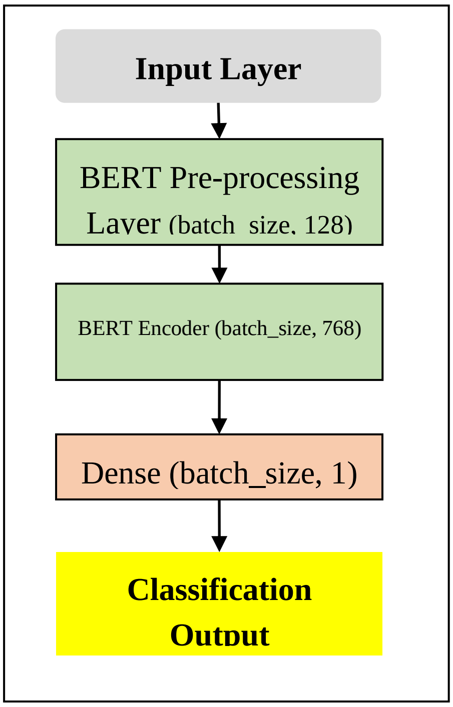
Input Layer
BERT Pre-processing Layer (batch_size, 128)
BERT Encoder (batch_size, 768)
Dense (batch_size, 1)
Classification Output

Fig 2: Standard BERT

Though the BERT architecture uses the transformers as its core structure, it differs in the other transformers where the BERT uses contextual and pre-trained embedding-layers and other architectures use non-contextual layers. According to Lee te al., (2021) to examine the codes and static data, BERT is found to be effective, significant and reliable than other transformers for textual data. By increasing a transformer's size, it is presumed by developers that the performance of a ML model is increased (Press, 2020) apart from DistilBERT. Thus, based on the literatures reviewed and analysed, the current research adopts the BERT architecture by increasing the size to heighten the performance as an advanced BERT model with additional layers.

3.1 Proposed architecture

The proposed architecture includes basic layers of BERT (transformer blocks, attention heads and parameters) with addition of two dropout layers and additional dense layer. The current research classifies the identified web vulnerabilities using the advanced BERT model and hence a classification layer is also included (refer to figure 3).

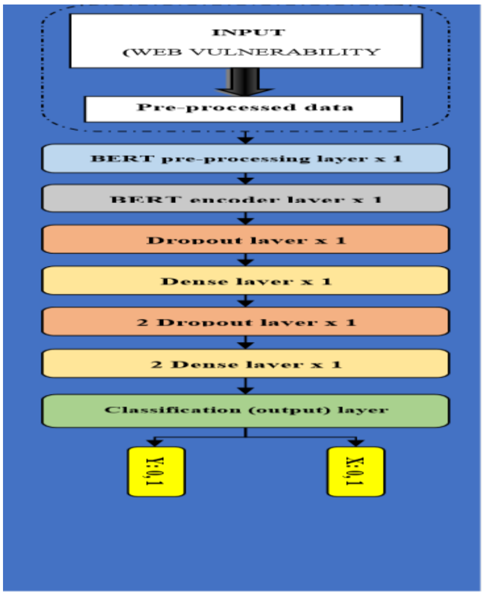

Fig 3: Developed advanced BERT model (ADBERT)

The architecture of the developed model functions as eight steps: initial step is to acquire the inputs (strings/texts) and processing them into pre-processed datasets. In the second stage the pre-processed datasets are passed via the BERT processing layer the raw texts are processed as numeric inputs for further processing. In this research the labels are defined as '0' and '1' where the '0' includes non-vulnerabilities and '1' includes vulnerabilities in the identified codes processed. In between the classification and processing layers, the model includes two dropout and dense layers, respectively.

3.2 Proposed algorithm and techniques

The machine learning includes different techniques, approaches and algorithms. According to the necessity and purpose, researchers adopt the algorithms. There are different types of algorithms in machine learning, namely:

a) Regression (polynomial and linear), random forest (RF), classification (support vector machine (SVM), tress, logistic regression, kNN, Naïve-Bayes) under the supervised ML;

b) Hidden-Markev model, association analysis and clustering (k-means, singular value decomposition (SVD) and principal component analysis (PCA)) under the un-supervised ML.

The current research adopts the transformer model which falls under the self-supervised category and thus the model uses the NLP application where algorithms like, SVM, conditional random-field, Bayesian networks, maximum entropy, neural networks (or deep learning) are adopted. In classification approach of a transformer model, the text categorization is done by methods namely: vector semantic, sequence-labeling, word embedding, text classification, probabilistic-language model and speech recognition. The current research model adopts the text classification of NLP for the research purpose.

3.2.1 Classification method adopted: Text classification in NLP assigns the set-of predefined categories for the open-ended texts in machine learning. It is used for structuring, organizing and categorizing the textual contents into sub-words or numerals as labels. These texts are acquired from files of different types like medical data, legal documents, coding, studies of various genres, and more mainly available through web or internet as resource.

3.2.2 Benefits of adopted-algorithm: There are different pros and cons in each algorithm adopted in the machine learning. The NLP has benefits namely: the implementation is easier, less costly than human intervention, execution and computing is faster than manual execution. Similarly, the AdamW as the optimizer algorithm is used by researchers since it is efficient than Adam, overcomes the issues in Adam algorithm like less memory, fast computing and convergence and lastly adaptive learning-rates.

3.3 Statistical approach and software used

3.3.1 Optimizer: The AdamW as the optimization algorithm is adopted here. It is a *stochastic optimization* method. AdamW modifies a typical weight decay implementation of Adam, towards combating the flaws and issues by decoupling the gradient updates' weight decay. The developed advanced BERT model is optimized using the optimizer algorithm where the performance and accuracy is adjusted as per the expectations. The formula used for the Adam regularization with the weight decay *wd* with *t* as time is represented as:

$$f_t = \forall f(\emptyset_t) + wd_t \emptyset_t$$

- whereas, the gradient update in AdamW with adjusted weight decay is represented as:

$$\emptyset_{t+1,k} = \emptyset_{t,k} - \lambda\left(\frac{1}{\sqrt{\vartheta_t + \varepsilon}} . \hat{\eta}_t + wd_{t,k}\emptyset_{t,k}\right), \nabla t$$

... (1)

3.3.2 Cross entropy for loss estimation: The loss of the model is estimated using the loss function in machine learning. This technique is also known among the researchers as 'log loss' method. The total number of classes in this research is represented as *C* the rows used for the dataset in rows are represented as *R*. The formula used to estimate the loss of the model is:

$$loss = -\frac{1}{R}\sum_m^R \sum_n^C x_{mn} \log(p_{mn})$$

..(2)

3.3.3 Performance evaluation metric: The model's performance is measured using the binary accuracy. It is calculated using the performance evaluation metrics. The accuracy formula is represented as:

$$Acc = \frac{TruePos + TrueNeg}{True + Pos + TrueNeg + FalsePos + FalseNeg} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textbf{(3)}$$

The positive values are the estimated and obtained true outcomes, whereas the negatives are unexpected outcomes. Thus the predicted and actually obtained true values are matched and weighed for accuracy estimation.

3.3.4 Software: The study uses 'python language' as the application. It is used for developing the model through in-built algorithms. Python is open-source and easy to understand. The BERT is originally developed by the Google-AI (Dong et al., 2020). In this research the advanced BERT is developed using the NLP tasks and text classification to identify and classify the web vulnerabilities. The model is trained in python to label the texts as numbers, though identified vulnerabilities in the SQLI and XSS datasets into '0' and '1' classifications. The python library for the BERT used is *pytorch*. The pre-trained BERT model is imported using the python script where the embeddings of the contextualized words are converted into numeric values as labels for classification. The in-built deep learning algorithms in python are used for classification.

3.4 Datasets

BERT model for detecting the vulnerability of web using two set-of diversified datasets, namely: SQLInjection vulnerability (SQLI) and Cross Site Scripting vulnerability (XSS).The datasets for both SQLInjection and Cross site scripting (XSS) is adopted from the databank "*kaggle*" as the source. The XSS datasets are acquired from the study by Shah (2020) and the SQLI datasets are acquired from study by Shah (2021).

SQLI: The SQLI dataset includes pre-processed and cleansed columns, raw SQLI attacks and different website based benign attacks. The databank has three .csv files and seven columns (five string and two integer columns). Using these columns the data as input is used for vulnerability detection. The columns include three columns where, the first column is of sentences as data scripts of 30873 unique values as datasets, columns two and three are labels. All three columns are of string values.

XSS: The cross site scripting has been used as the second dataset. It includes both benign and attack datasets. It is pre-processed and readily available to be used without pre-processing or cleansing. It has three columns like SQLI data where second column is a string with columns two and three as integer values. The third column represents the target integer. The first column includes 13.7thousand datasets.

Data loading: In the first step the data loaded are three CSV files that are merged as one file for the detection model to perform the tasks. The second step is to remove the rows in which the values of the label are not of 0 and 1. In this datasets, the NaN (Not-a-Number) values are dropped in the third step as this model and approach has no valid-way of filling-in the NAN values. In the fourth step of data loading, data repetition is removed where it is considered as insignificant sentences. Finally in the fifth step, the datasets are split into the ratios of 80:20 for training and testing where the whole data-frames are shuffled and divided.

Data labelling: The datasets are processed and classified in the output layer. The classification of input includes two sets of data predictions based folders with label 'x' and 'y' as annotations. The identified and predicted vulnerabilities are classified by the models as '0' and '1' where 0 represents there are no incorrect strings whereas 1 represents that the strings has vulnerabilities. Based on these classifications of the detected web vulnerabilities, the models' performances are estimated using the metric evaluation.

4. Results and Analyses

The results of the model's performance are obtained and represented through tables, graphs and charts.

4.1 Results

The results are divided into two sections, namely: base BERT model and advanced BERT (ADBERT) model analyses, respectively. The epoch runs, accuracy and loss evaluations are performed on the datasets and the results are portrayed through graphical and tabular forms.

The model was optimized by fine-tuning the parameter 'learning rate' (lr) from 1e-5 to 3e-5. The epoch was maintained as 5. The analyses are:

4.1 Analysis of basic BERT model:

i. XSS analysis:

Epoch	Time	Training-loss	Training-accuracy	Validation-loss	Validation-accuracy
1	112s 188ms	0.6798	0.5577	0.5926	0.6712
2	101s 186ms	0.5407	0.7395	0.5399	0.7326
3	111s 203ms	0.5028	0.7819	0.5137	0.7637
4	111s 203ms	0.4839	0.7972	0.5002	0.7830
5	102s 186ms	0.4755	0.8437	0.4958	0.7866

Table 2: Epoch values for performance of the classification test-model

The test model obtained the loss as 0.4854 at the time of 11s 162ms/step with accuracy of 0.7821 (78%). At 12s 164ms/step, the model obtained the validation accuracy of 0.7821 and training accuracy of 0.8437 (84%), and it remained the same as previous value (refer table 2). It can be understood from the results that the base model achieved 80% training accuracy in the classification of XSS analysis. The same is represented through graphical data (refer to figures 4 and 5). The loss of the base BERT model is estimated as value 0.48.

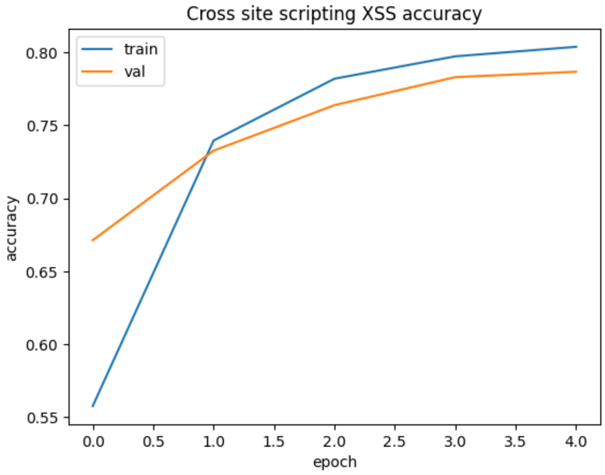

Fig 4: XSS accuracy – base model

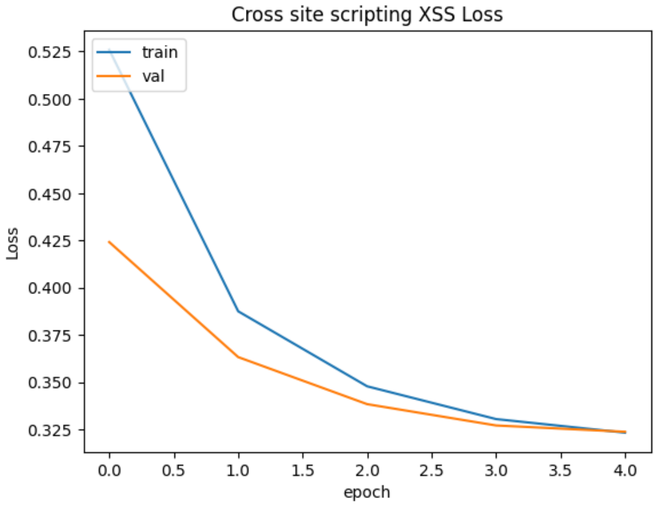

Fig 5: XSS loss – base model

Thus the cross site scripting accuracy analysis for classification of web vulnerability in the BERT model is observed as 80%.

ii. SQLI analysis:

The table 2 shows the base model's SQLI loss and accuracy values.

Epoch	Time	Training -loss	Training - accuracy	Validation -loss	Validation -accuracy
1	305s 387ms	0.5259	0.7482	0.4241	0.8372
2	295s 386ms	0.3875	0.8510	0.3633	0.8578
3	294s 385ms	0.3478	0.8640	0.3384	0.8679
4	296s 386ms	0.3305	0.8686	0.3271	0.8725
5	294s 385ms	0.3233	0.8707	0.3237	0.8715

Table 3: Epoch table for SQLI test set

From table 3 it's observed that, at 57s18ms/step the loss is 0.6507 with accuracy of 0.8673. However, later at time 47s15ms/step the accuracy of the SQLInjection model remained the same as 0.8673 (87%). The loss of the SQLI in BERT is estimated as value 0.32.

The graph (refer to figures 6 & 7) represents the accuracy and loss estimation, respectively.

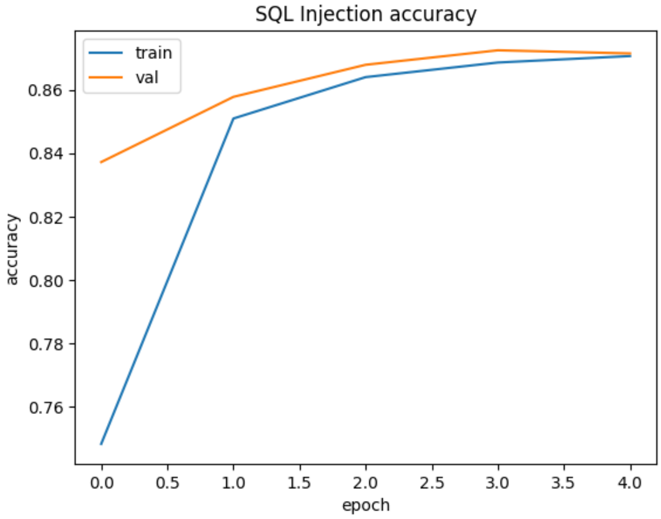

Fig 6: SQLI accuracy – base model

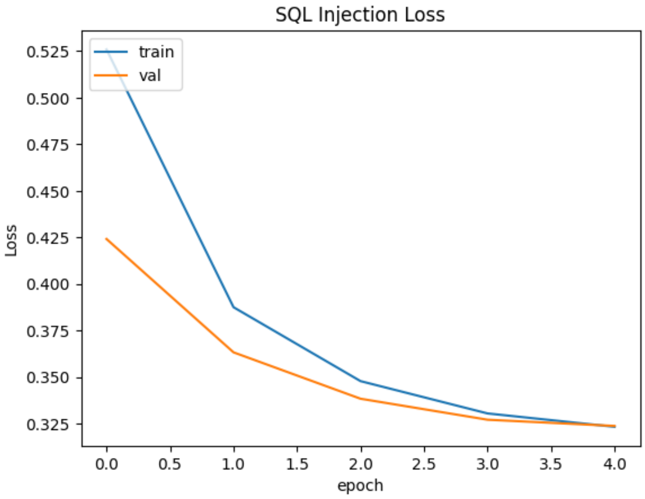

Fig 7: SQLI loss – base model

Thus the SQLInjection accuracy analysis for classification of web vulnerability in the BERT model is observed as 87%.

4.2 Analysis of advanced BERT (ADBERT) model:

i. Cross site scripting analysis:

Similar to the BERT analysis the ADBERT analysis also has same datasets and lr (learning rate). The model runs through 5epochs with initial LR as 1e-5. Later by optimization algorithm, the learning rate is fine tuned as 3e-5.

The XSS datasets are analysed and the results are:

Epoch	Time (s.ms/step)	Training-loss	Training-accuracy	Validation-loss	Validation-accuracy
1	722s 921ms	0.0494	0.9819	0.0133	0.9964
2	703s 919ms	0.0063	0.9987	0.0058	0.9990
3	703s 919ms	0.0028	0.9993	0.0069	0.9993
4	703s 918ms	0.0006.2436	0.9998	0.0066	0.9993
5	704s 920ms	0.0001.8842	0.9999	0.0069	0.9993

Table 4: Advanced BERT epoch table for XSS test set

From table 4 and analysis it is witnessed that, the ADBERT's loss estimated is of value 0.0033 at 58s 19ms. Similarly, the accuracy achieved is of value 0.9993 (99%) at 60s 19ms for the XSS classification.

The figures 8 and 9 represent the graph plots of the training and validation accuracy and loss estimations of ADBERT model with XSS datasets, respectively.

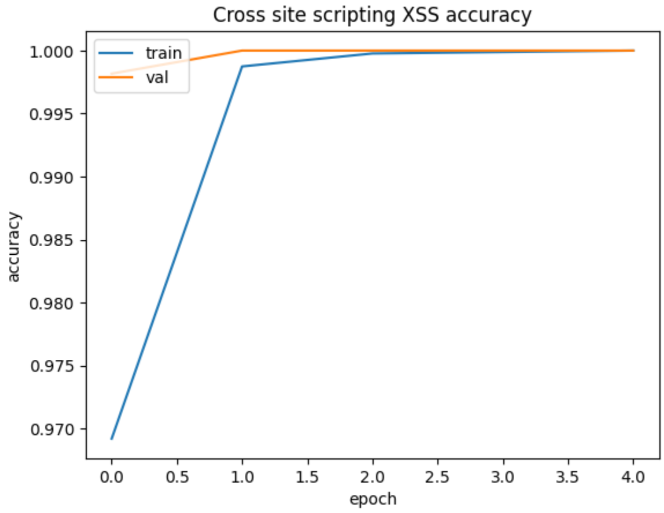

Fig 8: XSS accuracy – Advanced model

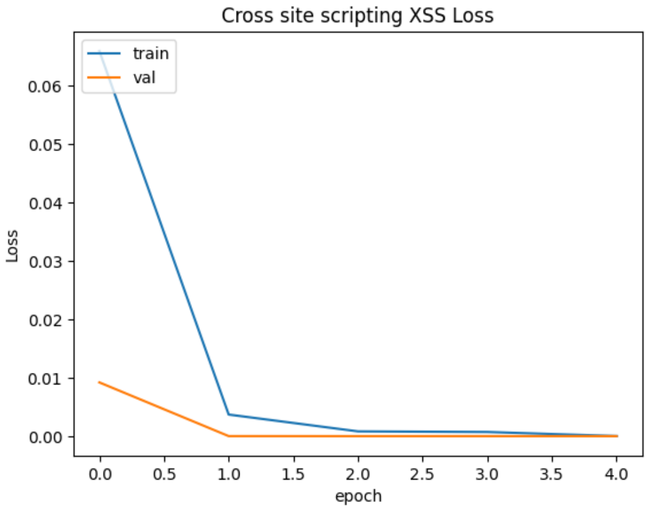

Fig 9: XSS loss – advanced model

The XSS accuracy analysis for classification of web vulnerability in the ADBERT model is observed as 99%.

ii. SQLI analysis:

The SQLInjection analysis of the ADBERT model is calculated as:

Epoch	Time (s.ms/step)	Training-loss	Training-accuracy	Validation-loss	Validation-accuracy
1	284s 486ms	0.0659	0.9692	0.0092	0.9982
2	263s 482ms	0.0037	0.9987	1.2315e-05	1.0000
3	263s 482ms	0.0008.2420	0.9998	5.6180e-06	1.0000
4	263s 482ms	0.0007.1982	0.9999	4.0659e-06	1.0000
5	263s 482ms	0.0001.8141	1.0000	3.1502e-06	1.0000

Table 5: Advanced BERT epoch table for SQLI test set

The table 5 shows loss value of the SQLI datasets in the ADBERT model is observed at 5th epoch with value of 0. Similarly, the accuracy was obtained as 100%. Based on the table values the accuracy and loss values are represented through graphical charts (refer to figures 10 and 11).

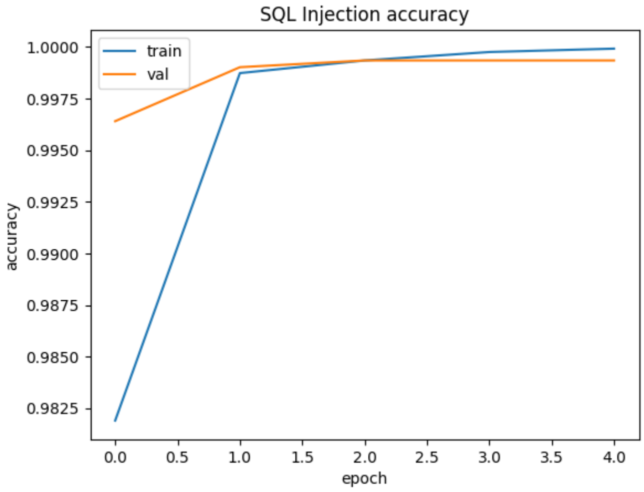

Fig 10: SQLI accuracy – advanced model

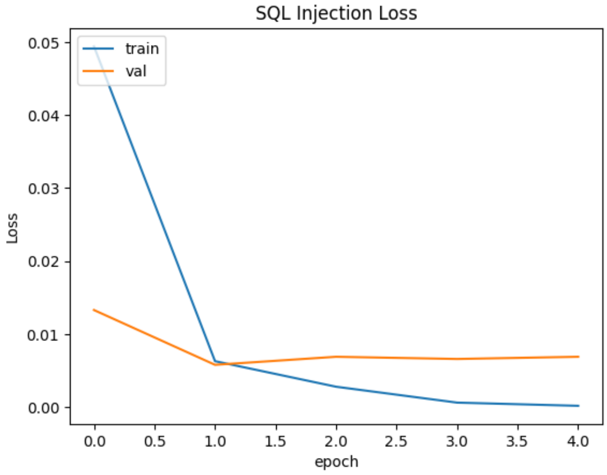

Fig 11: SQLI loss – advanced model

The SQLInjection accuracy analysis for classification of web vulnerability in the ADBERT model is observed as 100%.

iii. Performance analysis of advanced BERT:

The performance of the developed model (advanced BERT) is evaluated using the evaluation metrics. Both datasets SQLI and XSS are adopted to analyse the model's performance. Through training, BERT and ADBERT models predicted and identified web vulnerabilities however the model's performance for accuracy is evaluated using the binary accuracy metric evaluation. The performance metric results of XSS and SQLI classifications are:

- ***XSS classification performance analysis***

	Precision	Recall	F1-score	Support
0	0.98	0.98	0.98	1926
1	0.97	0.96	0.96	1134
Accuracy			0.97	3060
macro avg	0.97	0.97	0.97	3060
weighted avg	0.97	0.97	0.97	3060

Table 6: Classification performance analysis on XSS

- *SQLI classification performance analysis*

	Precision	Recall	F1-score	Support
0	0.98	0.97	0.97	371
1	0.98	0.99	0.99	721
Accuracy			0.98	1092
macro avg	0.98	0.98	0.98	1092
weighted avg	0.98	0.98	0.98	1092

Table 7: Classification performance analysis on SQLI

From tables6,it is observed that the performance of the XSS classification is significant with accuracy and precision as 97% in the advanced BERT model. Similarly the SQLI performance is observed (refer to table 7) as 98%. Thus proving the research purpose that, the developed advanced BERT model is more accurate with minimal loss than the base BERT model in web vulnerability detection and classification.

4.2 Performance and comparative analyses of the BERT models

Though the epoch values and accuracy with losses are obtained, the model's performance through evaluation metrics is estimated (refer to figure 11).

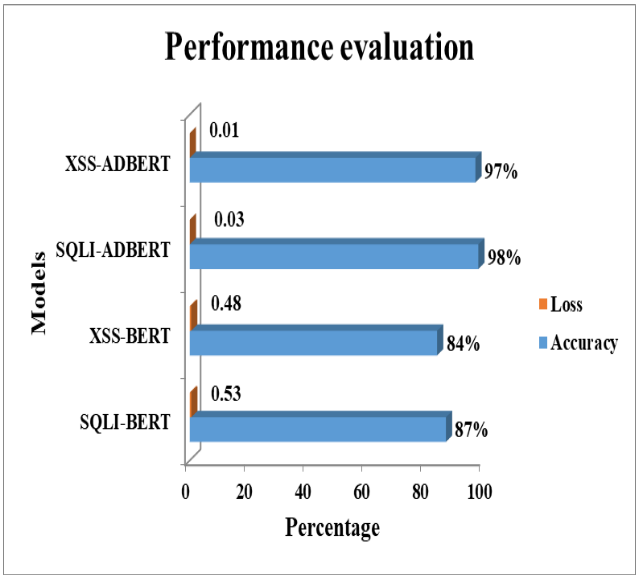

Fig 12: Performance evaluation

The findings of the analyses are represented as:

- SQLI accuracy is observed to be 87% and loss value of .53 at 5th epoch in the base BERT model classification and
  XSS accuracy is observed to be 84% and loss value of .48 at 5th epoch in the base BERT model classification.

- SQLI accuracy is observed to be 98% and loss value of .03 in the base ADBERT model classification and
- XSS accuracy is observed to be 97% with the loss value of .01 in the base ADBERT model classification.

The existing models on detecting the web vulnerabilities using the SQLInjection and XSS attacks has been attempted by researchers. However the techniques, algorithm, statistical approaches and application used are different. The table 8 shows the comparison of such studies attempted to examine, detect and predict the web vulnerabilities.

S. No	Author	Year	Model	Accuracy
1	Rawat et al.,	2012	SVM with SQLIA datasets	96.4%
2	Hasan et al.,	2019	SQLIA with Graphical user interface (GUI)	94%
3	Lu et al.,	2023	Semantic learning-based SQLIA	94%
4	Falor et al.,	2022	SQLIA with CNN-SVM model	95%
5	Dawadi et al.,	2023	DDoS, SQLIA and XSS datasets with Web-app using Long Short-Term Memory approach	97%
6	Wong and Luo	2020	BERT with Support Vector Classification (SVC) model	93%
7	Yang et al.,	2022	BERT with Advanced Persistent Threats (APTs)	90%
8	Proposed ADBERT model	2023	SQLI and XSS datasets	SQLI: 98% XSS: 97%

Table 8: Predictive ML models - Web vulnerabilities using XSS and SQLI datasets

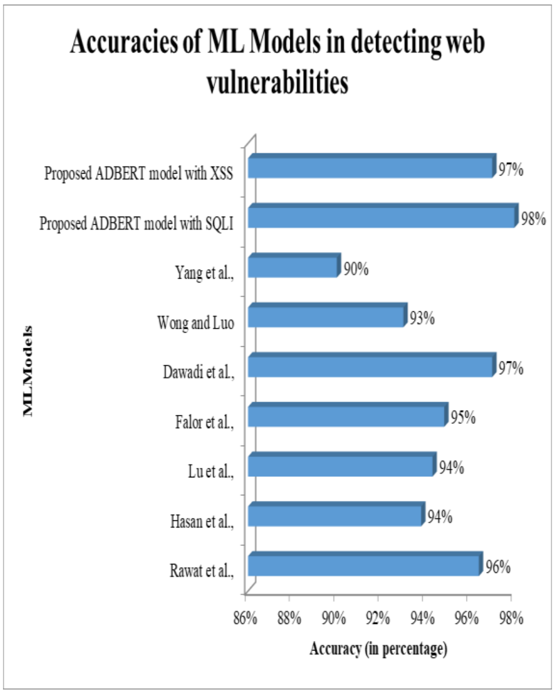

Fig 13: Comparative analysis

From figure 13 it is understood that the developed model ADBERT in predicting and classifying the web vulnerabilities with more than one dataset is significant and effective than the earlier models. The model developed is different from the basic BERT model with advanced layers and thus the name, ADBERT. The existing studies used SVM approach, random forest algorithm, CNN and ANN as the model architecture. The current research developed a ML based ADBERT model and attained 98% in SQLI detection and classification, whereas 97% in XSS detection and classification. Other models either used a hybrid approach for more than one vulnerability detection and classification, or, a single algorithm to predict and classify multiple datasets. The developed model is the first-of-its kind, since it has adapted BERT with multiple datasets for both prediction and classification. The model developed achieved higher accuracy with minimal loss, than existing BERT models.

5. Discussion and Conclusion

5.1 Discussion

The BERT as transformer uses only the encoder since it transforms the scripts and texts into numerals. The decoder is not used in BERT models. The BERT models are majorly used for text translation and conversion of textual contents into numeric contents through labeling. To identify and to classify the web vulnerabilities BERT models have been used. In this research the datasets used are the SQLInjections (SQLI) and the cross site scripting (XSS) based attacks and benign data. The relevant and existing researches have used different ML approaches namely, SVM, RF, Naïve Bayes, and more, with different architectures like ANN, RNN and CNN to identify and classify the textual contents.

Authors Bogale and Tamiru (2021) used the SHA512 to classify the web vulnerabilities using XSS and SQLI attacks and achieved more than 80% accuracy rate. Similarly, authors Johari and Sharma (2012) developed a hybrid model of AES with RSA based CNN architecture. The model analysed the datasets XSS and SQLIA and achieved approximately 90% accuracy. Study by author Ross (2018) developed an ANN architecture based model with four different algorithms J48, SVM, RF and JRip using the webapp. The author used correlated and datiphy based SQLI datasets and achieved 97% as accuracy. Though the study compared four different algorithms, the author developed the models with single classification for each model and it didn't focus upon multiple classifications. Sukhanand and Sharma in 2017 developed an evolutionary fuzzing based interface model by using the datasets SQLI and XSS and found that accuracy was less than they expected. Authors Rawat et al., (2012) developed a SVM with CNN model by using SQLIA datasets and achieved 96.4% as accuracy.

The studies and researches have either focused on single dataset or a hybrid model to achieve higher accuracy. There were no studies in textual classification using the BERT model with SQLInjection and XSS datasets. Hence, the current research focused on developing a transformer model. The research is the first-of-its-kind to develop a BERT transformer as architecture towards identifying and classifying the web vulnerabilities using the SQLI and XSS datasets. The developed model used additional layers with the standard layers of BERT to increase its performance. It is evident that the study is a success where transformer model achieved higher accuracy with multiple classification layers.

5.2 Conclusion

The study developed an advanced BERT model with additional dense and dropout layers to the standard layer. The BERT is a pre-trained transformer-encoder architecture that uses NLP application. Datasets used are obtained from 'kaggle' as resource. Two types of datasets are used to measure the web vulnerabilities namely SQLInjections (attacks and benign) and cross site scripting. The model is developed using the python language; to estimate the model's loss, binary cross entropy as loss estimation function is used. The accuracy is measured through the performance metric evaluation method.

The study found that, by increasing the BERT's layers, the performance of the BERT model also increased efficiently and significantly. The advanced model developed (ADBERT) produced 98% accuracy in SQLI dataset and 97% accuracy in XSS dataset whereas the standard BERT model produced less than 90% in both dataset classifications. Thus it is proved through the current research that, by optimizing and adding layers of BERT model, the accuracy can be increased effectively. The NLP application with text classification in this research is proved to be more reliable 'tasks' in classifying the texts and converting them into numbers. Data handling and complex issue of data conversion (text to numerals) was also handled well by the transformer model. The results obtained shows that web vulnerabilities identified by the ADBERT model shows significant outcomes in multiple classification.

5.3 Future implications

The web vulnerabilities caused by the third parties via attacks and injections have increased as the technology advancements bloomed. The common web vulnerabilities like injection flaws and cross site scripting have been focused in this research. In future different set of attacks using the ADBERT model will be examined. There are various web vulnerabilities to focus in the future like: broken authentication, insecure direct object references, sensitive data exposures, security misconfiguration, missing function-level access-control, known vulnerabilities based component attacks, cross-site request forgery (CSRF), un-validated forwards and redirects and more.

References:

[1] Alghawazi. M, Alghazzawi. D and Alarifi. S., (2022), "Detection of SQL Injection Attack Using Machine Learning Techniques: A Systematic Literature Review", *Journal of Cyber Security and Privacy,***2**: 764-777.

[2] Azman. M.A, Marhusin. M.F and Sulaiman. R, (2021), "Machine Learning-Based Technique to Detect SQL Injection Attack", *Journal of Computer Science,* **17(3)**: 296-303.

[3] Barde. S.S, (2020), "Cross Site Scripting detection using Random Forest Bagging and Dataset Ensemble Modelling", MSC thesis submitted to National College of Ireland – School of Computing, 1-19, August 2020.

[4] Chen. D, Yan. Q, Wu. C and Zhao. J, (2020), "SQL Injection Attack Detection and Prevention Techniques Using Deep Learning", *Journal of Physics: Conference Series,***1757(2021):** 1-8.

[5] Johari. R and Sharma. P, (2012), "A Survey On Web Application Vulnerabilities (SQLIA, XSS) Exploitation and Security Engine for SQL Injection", In: *2012 International Conference on Communication Systems and Network Technologies (IEEE – Computer Society),* **2012**: 453-458.

[6] Kumar, R. (2011), “Mitigating the authentication vulnerabilities in Web applications through security requirements,” *Information and Communication Technologies (ICT)*,**60**: 651-663.

[7] Lee. N.Q.K, Ho. Q-T, Nguyen. T-T-D and Ou. Y-Y, (2021), “A transformer architecture based on BERT and 2D convolutional neural network to identify DNA enhancers from sequence information”, *Briefings in Bioinformatics*, **2021**: 1-22.

[8] Press. O, Smith. N.A and Levy. O, (2020), “Improving Transformer Models by Reordering their Sub-layers”, In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, **2020**: 2996-3005. July 5 - 10, 2020.

[9] Ross, K., (2018), “SQL Injection Detection Using Machine Learning Techniques and Multiple Data Sources”, *Master's Projects*, **650**: 1-28. Available at https://scholarworks.sjsu.edu/etd_projects/650

[10] Sukhanand. S and Sharma. P, (2017), “A Review Paper on SQL Injection and Cross Site Scripting Vulnerabilities”, *International Journal of Creative Research Thoughts (IJCRT)*, **5(4)**: 3721-3724.

[11] Kumar. A and Binu. S, (2018), “Proposed method for SQL injection detection and its prevention”, *International Journal of Engineering and Technology*, **7:** 213.

[12] Rahman. T.F.A, Buja. A.G, Abd. Kand Ali. F.M, (2017),“SQL Injection Attack Scanner Using Boyer-Moore String Matching Algorithm”,*JCP*, **12(2):** 183-189.

[13] Fang. Y, Peng. J, Liu. L,and Huang. C,(2018), “WOVSQLI: Detection of SQL Injection Behaviors Using Word Vector and LSTM”, In:*Proceedings of the 2nd International Conference on Cryptography, Security and Privacy*, Guiyang, China, 16 March 2018; ACM: Rochester, NY, USA. 170-174.

[14] Gong. X, Zhou. Y, Bi. Y, He. M, Sheng. S, Qiu. H, He. R, and Lu. J,(2019), “EstimatingWeb Attack Detection via Model Uncertainty from Inaccurate Annotation”, In:*Proceedings of the 2019 6th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)*, Paris, France, June 21-23 June.*IEEE: Piscataway*, NJ, USA. 53-58.

[15] Abdulmalik. Y, (2021), “An Improved SQL Injection Attack Detection Model Using Machine Learning Techniques”, International Journal of Innovation in Computers**11**: 53-57.

[16] Farooq. U,(2021), “Ensemble Machine Learning Approaches for Detection of SQL Injection Attack”,*Teh. Glas*, **2021(15):** 112-120.

[17] Dong. J, He. F, Guo. Y, et al.,(2020), “A Commodity Review Sentiment Analysis Based on BERT-CNN Model”, In: *2020 5th International Conference on Computer and Communication Systems (ICCCS)*, 2020: 143-147.

[18] Bisht. P, Madhusudan. P and Venkatakrishnan. V.N, (2010), “CANDID: Dynamic Candidate Evaluations for Automatic Prevention of SQL Injection Attacks”, *ACM Trans. Inf. Syst. Secur.*,**13(2)**:1-39.

[19] Indrani Balasundaram, E.Ramaraj “An Authentication Scheme for Preventing SQL Injection Attack Using Hybrid Encryption (PSQLIA-HBE)”, *European Journal of Scientific Research*, **53(3)**: 359-368.

[20] Srivastava. N, Hinton. G, Krizhevsky. A, Sutskever.I and Salakhutdinov. R,(2014), “Dropout: A simple way to prevent neural networks from overfitting”, *Journal of Machine Learning Research,* **15**:1929-1958.

[21] Dawadi. B.R, Adhikari. B, and Srivastava. D.K, (2023), “Deep Learning Technique-Enabled Web Application Firewall for the Detection of Web Attacks”, *Sensors,* **23**: 2073.

[22] Yang. X, Peng. G, Li. Z, Lyu. Y, Liu. S and Li. C, (2022), “Research on entity recognition and alignment of APT attack based on Bert and BiLSTM-CRF”, *Journal of Communications,* **43(6)**: 58-70.

[23] Wong. H and Luo. T, (2020), “Man-in-the-Middle Attacks on MQTT-based IoT Using BERT Based Adversarial Message Generation”, In: KDD’20 Workshops (AIoT), August 24, SanDiego, CA. 1-6.

[24] Lu. D, Fei. J and Liu. L, (2023), “A Semantic Learning-Based SQL Injection Attack Detection Technology”, *Electronics*. **12(6)**:1344.

[25] Hasan. M, Balbahaith. Z and Tarique. M, (2019), “Detection of SQL Injection Attacks: A Machine Learning Approach," In: *2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, Ras Al Khaimah, United Arab Emirates, 1-6.

[26] Falor. A, Hirani. M, Vedant. H, Mehta. P, and Krishnan. D, (2022), “A Deep Learning Approach for Detection of SQL Injection Attacks Using Convolutional Neural Networks”, In: *Gupta, D., Polkowski, Z., Khanna, A., Bhattacharyya, S., Castillo, O. (eds) Proceedings of Data Analytics and Management. Lecture Notes on Data Engineering and Communications Technologies*, Vol. 91. Springer, Singapore.

[27] Shah. S.S.H, (2020),*“Cross site scripting XSS dataset for deep learning”*, Available athttps://www.kaggle.com/datasets/syedsaqlainhussain/cross-site-scripting-xss-dataset-for-deep-learning

[28] Shah. S.S.H, (2021), *“SQL Injection dataset”.* Available athttps://www.kaggle.com/datasets/syedsaqlainhussain/sql-injection-dataset

[29] PreciseSecurity.com (2019), *“Cross-Site Scripting (XSS) Makes Nearly 40% of All Cyber Attacks in 2019”,* Available at

https://www.precisesecurity.com/articles/cross-site-scripting-xss-makes-nearly-40-of-all-cyber-attacks-in-2019/

[30] Statista (2020), Vailshery. L.S, *"Global web application critical vulnerability taxonomy 2022".*In Technology and Telecommunications article. Available at https://www.statista.com/statistics/806081/worldwide-application-vulnerability-taxonomy/

[31] InvistiSecurity.com (2022), *"Report: 35% of educational institutions have a SQLI vulnerability".* Available at https://venturebeat.com/security/report-35-of-educational-institutions-have-a-sqli-vulnerability/

[32] Merritt. R, (2022), "*What Is a Transformer Model?*". Available at https://blogs.nvidia.com/blog/2022/03/25/what-is-a-transformer-model/#:~:text=A%20transformer%20model%20is%20a,25%2C%202022%20by%20Rick%20Merritt

[33] Application Defence Centre (ADC), 2015, "2015 Web Application Attack Report (WAAR)", available at http://www.imperva.com/docs/HII_Web_Application_Attack_Report_Ed6.pdf