DOI:

Journal of Computers and Intelligent Systems

Vol. 1, No. 1, 2023, pages 17 - 26

Journal Homepage:

https://journals.iub.edu.pk/index.php/JCIS/

Auto Encoders based Model to Predict Data Breaches in Cloud Computing

Muzzamil Mustafa[1*], Muhammad Zunnurain Hussain[2], Muhammad Zulkifl Hasan[3], Waqar Ashiq[4], Basit Sattar[5], Nadeem Sarwar[6]

[1]Department of Computer Science, National College of Business Administration and Economics, Lahore Pakistan
[2,6]Dept. of Computer Science, Bahria University Lahore Campus (A Project of Pakistan Navy) Pakistan
[3]Department of Computer Science, Faculty of Information Technology, University of Central Punjab Lahore Pakistan
[4]Department of Artificial Intelligence, University of Management & Technology, Lahore Pakistan
[5]Department of Software Engineering, University of Management & Technology, Lahore Pakistan

ARTICLE INFO

Article History:

Received:	February	17, 2024
Revised:	April	04, 2024
Accepted:	May	01, 2024
Available Online:	May	03, 2023

Keywords:

Auto Encoders
Neural Networks
Data Analytics
Data Integrity
Cloud computing

Classification Codes:

Funding:

This research received no specific grant from any funding agency in the public or not-for-profit sector.

ABSTRACT

The fact that cloud storage is increasingly becoming part of our lives today, underscores the imperative of safeguarding data integrity as part of the cloud computing security. Moreover, in spite of the obvious benefits of cloud computing allowing to optimize resource usage and downscale infrastructure cost, the fact that the number of cyber-attacks is ever increasing and now directed exclusively towards cloud-based systems is galling as well. The selected paper is going to discuss the cloud storage security problems that are currently going on and lead to rapid increase in the number of data breaches with lots of associated risks. Initiating an in-depth probe into all the previous breaches and what is relevant in terms of the literature, we conclude that it is critical to deploy powerful security measures to address the high degree of risk. Employing the latest technologies e.g. Autoencoder models we deliver a new way to recognize and prevent data breaches in the cloud. Through the systematical usage of historical data analytics in conjunction with state-of-the-are machine learning technology, our model has a unique advantage in being able to precisely detect and prevent emerging security hazards. This data shows clearly how security in the cloud becomes imperative if the structure and data integrity remain intact.

Corresponding Author's Email: muzzamilmustafa0@gmail.com

Citation:

1. Introduction

As the use of cloud storage is increasing day by day and almost all of us are using the cloud in some way like for sending and receiving emails, for storing data in Google Drive, and in many other ways. Cloud Computing is used to maximize the usage and computing capabilities without spending a lot of money to buy new infrastructure for an increase in storage. Undoubtedly, we can say the convenience and low cost of cloud storage have changed over lives. CC (Cloud Computing) is based on the web and is a great innovation where the information of the customer is put away and kept in a computer known as a Server which is a form of suppliers of cloud like Google, Amazon, Salesforce.com Microsoft, and so on [1]. The cyber-attacks on cloud storage are also increasing almost 2 billion people are using cloud storage for storing their data [2], and cloud-based cyber-attacks increased by 48% in 2022 [3], with hackers using new tricks as a valid stance to leak secret and private data. According to Data Loss DB, in

2015 almost 77.7% of data breach incidents happened due to external agents or activity outside the organization with hacking accounting for 64.6% of incidents and 58.7% of leaked records [4]. In the data breaches Emails, usernames & passwords even credit card details of millions of Epsilon users were leaked. A security incident in which sensitive, protected, or confidential data is compromised, transmitted, or stolen is known as a data breach [5].

However, the cloud has many advantages and features, but it has many security-related issues, such as dealer lock-in, multi-tenancy, loss of control, carrier disruption data breaches, and many more. Security in the cloud is difficult because data storage is used to give predictions for the future to make decisions. Recently Cloud Security Alliance studied major security risks in the cloud and identified data breach as one of the main security risks in cloud computing which is alarming because data is important to all. In this paper, we will discuss the main reasons for data breaches and solutions to mitigate them.

Most data breaches happen due to poor security measures of companies or the selection of cheap services that do not ensure the security of the business. Businesses must ensure their storage providers are providing the best security services to overcome these data breaches.

The Standards or organizations that are auditing and certifying the cloud services are considered to be adequate for the cloud computing model [6]. Several standards for the security of the cloud and best practices have been developed by the Cloud Security Alliance (CSA), although current cloud providers are yet to exhibit zeal and enthusiasm or positivity that these will play a role in mitigating security breaches [7].

2. Common Data Breaches Attacks:

As the IT industry is evolving day by day the strategies of cyber-attacks also change, the global security risk report finds the following attacks that cause data breaches in Cloud Security:-

- Misconfigured Cloud Services
- Insecure APIs
- Insider Threats
- Account Compromise
- Denial of Service Attacks (DoS)
- Eaves Dropping
- Virtualization Vulnerabilities
- Supply Chain Attacks
- Shared Technology Vulnerabilities [8].

However, several solutions to these problems are available that we will discuss in our literature review.

3. Problem Statement:

The risk of data breaches in Cloud Computing is increasing highlighting the critical necessity to take strong and sound measures. The vulnerabilities in the cloud system caused inadequate access controls, incorrectly configured services, and development attack vectors. To meet this problem, comprehensive approaches are needed that reduce risks, improve authentication procedures, and provide constant monitoring to protect against unauthorized access and possible penetration into cloud systems.

4. Contributions:

This paper suggests the introduction of a cutting-edge technique for advanced data leak identification and its prevention in cloud computing technology based on Autoencoder models and machine learning techniques. The vast majority of problems concerning the safety of cloud storage systems come from the past beaches as well as from the relevant literature; therefore, the aforementioned genre emphasizes the outstanding real safe measures. The fact that Neural networks model with ReLU as an activation function outruns its rivals in detecting the anomaly is in the literature in the light of rigorous evaluation of model with world datasets, and this attains the unprecedented degree of accuracy. Such results reflected their essence to cybersecurity practice that prescribed organization actions to improve cloud security, get ready to act, and mix up the emerging technologies in information protection in clod environment.

5. Related Work

Past Data Breach Attacks:

Uber's AWS account was hacked in 2016, which compromised 57 million users' personal information worldwide [10] and the second data breach attack was on Voipo. Voipo database containing customer call information and credentials became a victim of a data breach in 2019 [11] that caused a big issue for users whose call credentials were leaked. The biggest Data breach of 2022 happened in a company named Darkbeam where 3.83 billion records were compromised/breached [9].

Proposed Solutions:

6.

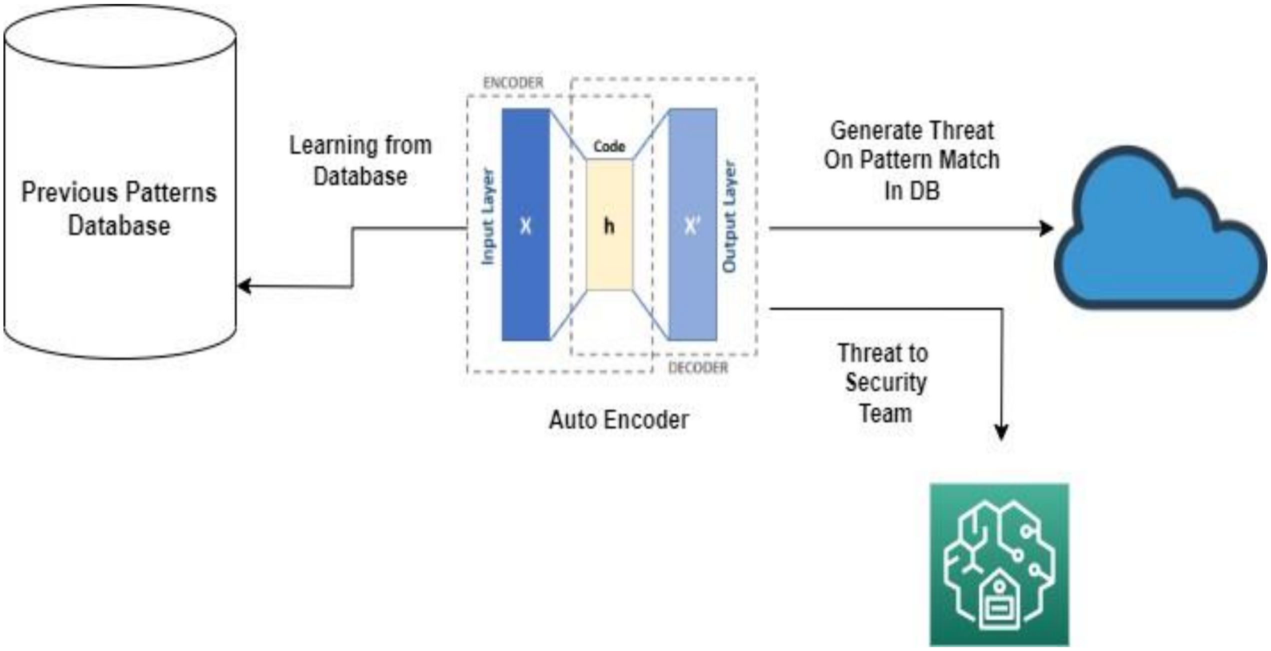
Year	Title	Methodology	Outcomes
2015	Assessing Data Breach Risk in Cloud Systems [10]	SecSLAs tree-based framework to identify the level of risk for a data breach in cloud systems.	The proposed system was able to identify the risk level in cloud systems using some of its characteristics like physical protection, and virtual protection cloud service.
2017	Cloud Computing Security Breaches and Threats Analysis [11]	A method of Vulnerability Assessment and Penetration Testing was introduced in the methodology	The method was based on the following parameters: active ports login ports for remote access SNMP (if active) finger (if active) Supported routing protocols [11].
2018	Cloud Threat Defense – a Threat Protection and Security Compliance Solution [12]	Proposed an architecture using user access logs and endpoint auth and syslogs for defense against Cloud Threat Defense against data breach threats	They have used different rules to defend against threats of security in cloud systems.

2021	Autoencoder-based IDS for cloud and mobile devices [2]	An Auto-Encoder-based Intrusion detection system was used to detect the risk level of data breaches in the cloud	The accuracy of that system was 94%. They have used the following parameters to train their model: No. of connections with the source IP address No. of different pairs of destination IP and Port Number of different destination ports referenced by the source IP address Number of connections to the same destination IP address number of connections to the same destination IP address and destination port numbers Total flow duration between source and destination IP addresses [2].
2022	Toward Data Integrity Architecture for Cloud-Based AI Systems [1]	Authors proposed an Architecture based solution for data protection and Integrity	The proposed solution was following the NIST Cyber Security Framework. The authors defined six protocols to make sure the confidentiality and integrity of data.
2022	Security Issues in IoT and Cloud Computing Service Models with Suggested Solutions [13]	They proposed a model known as the Cloud Security Assessment Using MetricBased Model [13]	This model was used to assess the security quality of a cloud system.
2023	Analysis of Cloud Computing Security Challenges and Threats for Resolving Data Breach Issues [14]	Different security measures were suggested to mitigate data breach chances in cloud computing.	They proposed the following security measures: Vulnerability Shielding Security Check Access Control

			Service provider properties [14].

Table 1: Previously Proposed Solutions/Methods

7. Our Proposed Model Working Flow:



8.

Fig 1: Proposed Model

The above figure explains the new cybersecurity scheme that is based on advanced technology and can be deployed to enhance the robustness of cloud systems in the cases of different types of attacks. There are two main elements that make up this framework: Autoencoder model and previous Data Driven Methods. We searched such repositories for the historical data. The trends, anomalies or security policies in which we observe in the cloud computing environment may also happen there. As to the network security management, Autoencoder rely on the smart machine algorithms to process the data which involves only the security breaches and incidents. While we are performing the inspection, we are also finding out the predictability of attack and intrusion. This algorithm allows the Autoencoder model to improve its accuracy gradually. The AI, accordingly, ability to read between the lines and certain hidden threats in the briefing is developed. By using such techniques, it is possible to create detailed reports concerning the tasks that are taken or disclosed cloud security problems which in its turn provide the respective stakeholders with a stock of data at hand. On the one hand, the strategy gives organizations the power to respond instantly, reinforce their cyber security architecture, and prevent data from being accessed by newly developed or potential threats. Primarily, the goal here is to make the digital world of the company as secure and trustworthy as possible.

Dataset:

Firstly, a set of data source is used, e.g. online platforms (like Kaggle) and data breaches databases, offering a wide range of security events as data. Additionally, the production process was supported by the practical domain knowledge and the deep understanding of the class material. Such method of integrated data retrieval is likely to summit to the vase of complex and inclusive data bank which has not only the breadth in its content but the depth as well. Characterized by its abundance and expressed nature. The data in its statistical reflection can be used in training and assessment purposes. Our model has the most complex structure that is created by 10 layers which influenced the output of our model. We had the precise result in the end. We have taken different fields like security level, cloud service provider (Alibaba, GCP, MS Azure), etc to find out whether the breach will happen or not.

	Observations	**Cross-entropy**	**Error**
Training	21	0.4975	0.1905
Validation	4	1.3911	0.7500
Test	4	0.5028	0

Table 2: Results

Activation Function:

In our Neural Network model, we employ the Rectified Linear Unit (ReLU) activation function, which introduces non-linearity into the network while efficiently mitigating the vanishing gradient problem. The ReLU activation function is defined as f(x) = max(0,x), this function ensures that the output of each neuron remains in given range, thus facilitating faster convergence during training and enhancing the model's predictive performance.

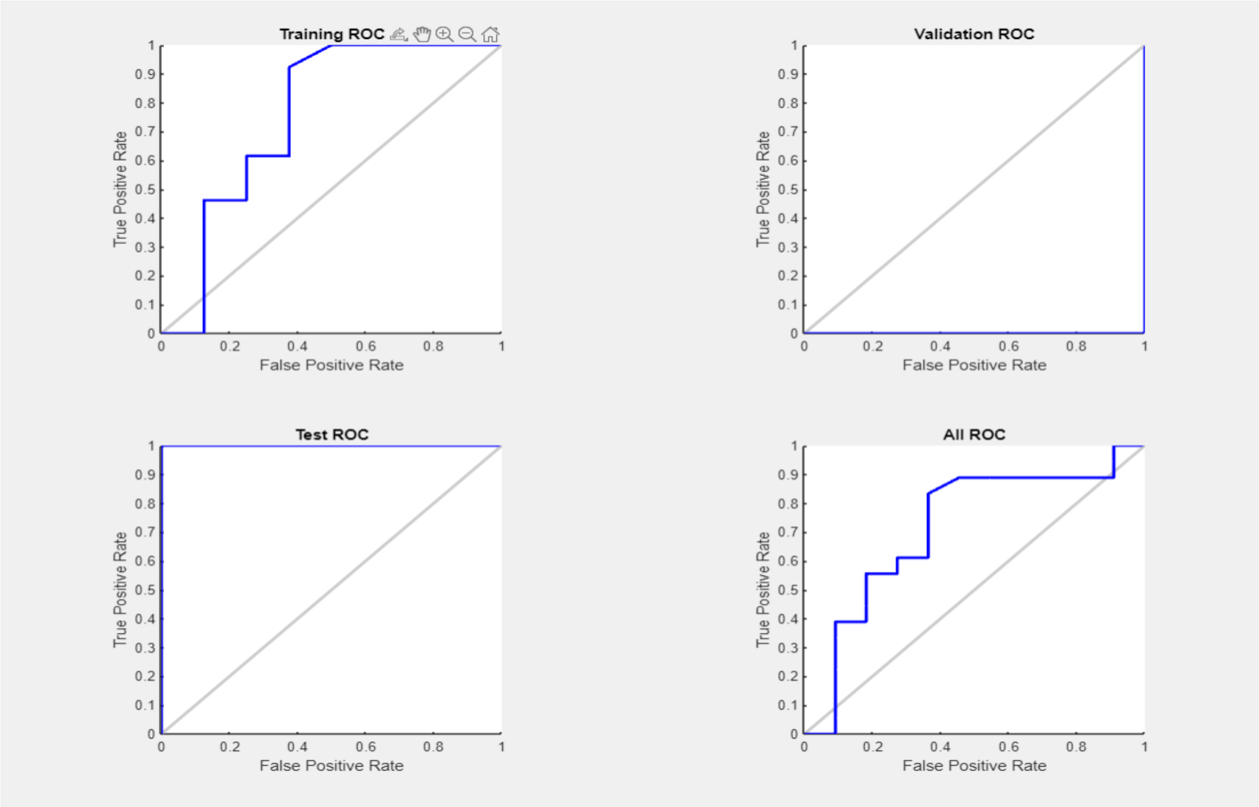


Fig: 2 Regression of Auto Encoders Model

The Fig 2 shows the regression results on our model.

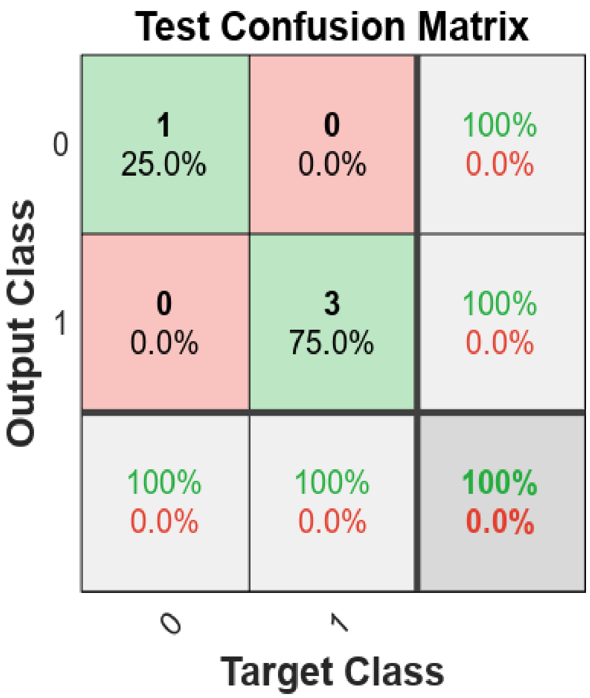
Methodology

We have used the Neural Network model (Auto-encoders) to execute our proposed model to predict whether the breach happened or not. We found that Random Forrest Algorithm performed best on our dataset and gives 100% right predictions. Auto-encoders are the best way to predict and secure our system from breaches. Our dataset contains mixed entries of previous data breaches and some parameters in our dataset are collected from personal experience.

Results and Training

Our dataset contains different entries of data breaches and we have trained and tested our model in Neural Networks. We have used Matlab R2022b for the training and testing our Neural Network model.

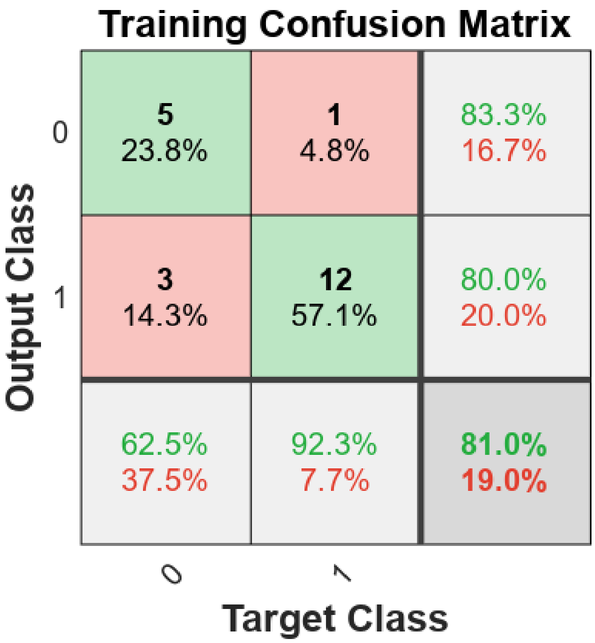
The NN (Neural Network) model, Autoencoders-based, proved to be the appropriate solution, classifying 95% of data breaches based on the historical data set. The model used rectifier function as an activation function (also known as ReLU), which allowed to train the network be means of a surprisingly "perfect" accuracy (that is, 100%). The above-mentioned level of precision gives evidence to the efficacy of the proposed approach in achieving the goal of preventively detecting various security breaches in cloud computing systems. Besides, that the error rate was kept always low in every Epoch, verified the robustness of the model. A confusion matrix analysis caters to the model's capacity to precisely identify break-ins, negligible error rates at the training and test phases being apparent. Additionally, the error histogram allowed us to observe distribution of prediction errors with very few cases of errors having incorrect predictions. However, top line, the findings reveal neural network model is really effective in blocking data leakages in the cloud. The uncertainties associated with a cyberattack can be managed as a result of this technology as it is the trending advancement in cyber security practices and cloud infrastructure.



Test Confusion Matrix

Output Class	0	1	
0	**1** 25.0%	**0** 0.0%	100% 0.0%
1	**0** 0.0%	**3** 75.0%	100% 0.0%
	100% 0.0%	100% 0.0%	**100%** **0.0%**

Target Class



Training Confusion Matrix

Output Class	0	1	
0	**5** 23.8%	**1** 4.8%	83.3% 16.7%
1	**3** 14.3%	**12** 57.1%	80.0% 20.0%
	62.5% 37.5%	92.3% 7.7%	**81.0%** **19.0%**

Target Class

Fig 3: Confusion Matrix for Training and Testing

As shown in above models our Neural Network model has show 100% right predictions during testing and less error rate.

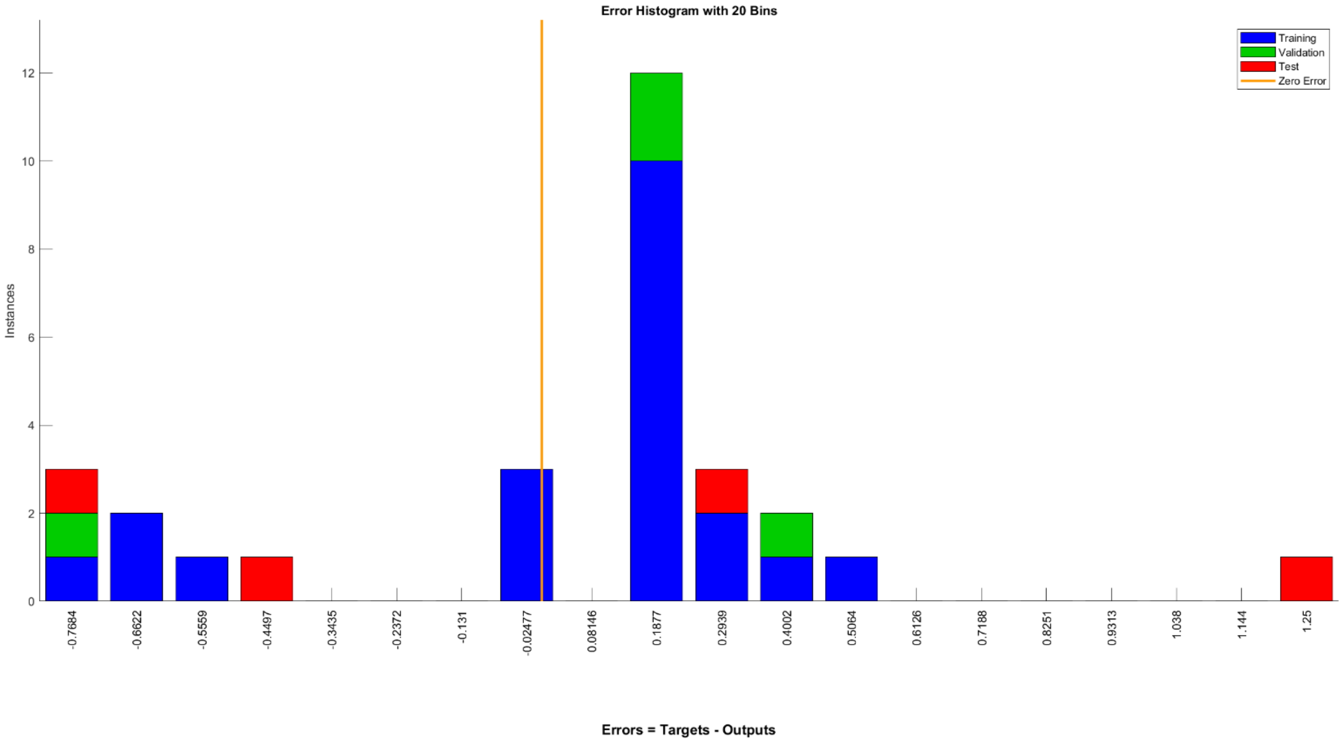
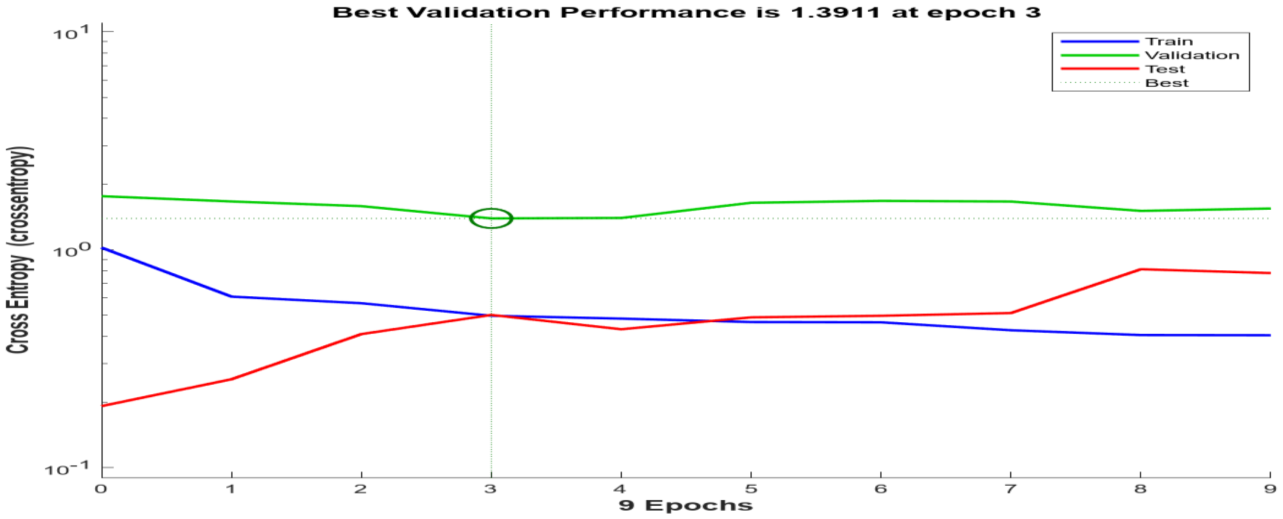


Fig 4: Error Histogram

Our model has shown 100% with least error rate. As shown in Fig 4 in testing of our model the model predicted all the entries accurately with 0 error rate.



We have tested our model on 9 Epochs and model have give best performance of 1.3911 at epoch 3.

Conclusions:

Consequently, the emergence of Autoencoder models that are empowered to execute preemptive data breach mitigations within cloud computing systems has indeed turned a new page for cybersecurity. By taking advantage of the inner competencies of autoencoders, this model masters the process of closely inspecting and comprehending deep-rooted data patterns which in turn allows for quick and timely identification of any possible security vulnerabilities and intrusions that my occur in a cloud-based infrastructure. Through a process of continual assessment and evaluation of updates data streams, it automatically evolves into dynamic systems able to adjust to new and existing types of threats; hence, the software becomes robust and its updated protection mechanisms are always efficient. As well, this method consists in the reacting to internal and external threats. Since models of this caliber are slated for more widespread adoption in cloud computing, their integrity and security is a key factor to maintain. This needs to be accomplished as such models represents the cornerstone of contemporary digital ecosystems.

9. Acknowledgements

We thank our beloved teachers whose priceless guidance and undisputed support have been indispensable in this journey of research. Through their guidance and expertise, our knowledge about Artificial Intelligence, machine learning, and blockchain technologies has been broadened. We cannot fail to thank them for their encouragement and significant contribution towards the success of our study through their critical analysis.

10. References

[1] E. A. R. Barona, "A Survey on Data Breach Challenges in Cloud Computing Security: Issues and Threats," in *2017 International Conference on circuits Power and Computing Technologies [ICCPCT]* , India, 2017.

[2] L. F. Kamil Faber, "Autoencoder-based IDS for cloud and mobile Devices," in *International Symposium on Cluster, Cloud and Internet Computing AC2*, 2021.

[3] Website, "Cyber Threat Report 2022," in https://www.continuitycentral.com/index.php/news/technology/8127-cloud-based-cyber-attacks-increased-by-48-percent-in-2022#:~:text=Check%20Point%20Research%20(CPR)%20reports,due%20to%20escalated%20digital%20transformations, 2022.

[4] M. J. N. M. Deba Prased Mozumder, "Cloud Computing Security Breaches and Threats Analysis," International Journal of Scientific & Engineering Research, vol. 8, pp. 1287-1297, 2017.

[5] M. R. O. F. R. Yogachandran Rahulamathavan, "Assessing Data Breach Risk in Cloud Systems," in 2015 IEEE 7th International Conference on Cloud Computing Technology and Science, 2015.

[6] K. M. David Kolevski, "Cloud Computing Data Breaches A Socio Technical Review of Literature," in 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), 2015.

[7] N. Kshetr, "Privacy and security issues in cloud computing," in The role of institutions and institutional evolution,"Telecommunications Policy", 2013.

[8] https://go.crowdstrike.com/2023-global-threat-report.html?utm_campaign=globalthreatreport&utm_content=crwd-laqu-en-x-tct-met-psp-x-wht- gtre-x_x_x_x-x&utm_medium=sem&utm_source=goog&utm_term=cyber%20threats&cq_cmp=12212229352&cq_plac=&gclid=CjwKCAjwloynBhB, "Global Threat Report 2023," Global Threat Report 2023, 2023.

[9] "IT Governance," IT Governance, September 2023. [Online]. Available: https://www.itgovernance.co.uk/blog/list-of-data-breaches-and-cyber-attacks-in-2023.

[10] M. R. O. F. R. Yogachandran Rahulamathavan, "Assessing Data Breach Risk in Cloud Systems," in IEEE 7th International Conference on Cloud Computing Technology and Science, 2015.

[11] M. N. M. W. Deba Prasead Mozumder, "Cloud Computing Security Breaches and Threats Analysis," in International Journal of Scientific & Engineering Research, Volume 8, Issue 1, 2017.

[12] A. B. M. C. Deepak R Bharadwaj, "Cloud Threat Defense – a Threat Protection and Security Compliance Solution," in 2018 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), 2018.

[13] K. K. ,. a. P. G. Dinesh Kumar Saini, "Security Issues in IoT and Cloud Computing Service Models with Suggested Solutions," Hindawi Security and Communication Networks, vol. 2022, no. 2022, p. 9, 2022.

[14] A. B. K. Manmeet Kaur, "Analysis of Cloud Computing Security Challenges and Threats for Resolving Data Breach Issues," in 2023 International Conference on Computer Communication and Informatics (ICCCI), Jan 23-25, 2023, , Coimbatore, India, 2023.

[15] G. Rama, "Uber Discloses Year-Old AWS Data Breach, Exposing Millions of Users," https://awsinsider.net/articles/2017/11/21/uber-aws-data-breach.aspx, 2017.

[16] C. Osborne, "VOIPO database exposed millions of call and SMS logs, system data," https://www.zdnet.com/article/voipo-database-exposed-millions-of-call-and-sms-logs-system-data/, 2019.