Journal of Analysis and Computation (JAC)
(An International Peer Reviewed Journal), www.ijaconline.com

Optimizing Cloud Security with Fusion Feature Selection Techniques and an Ensemble Classifier for Intrusion Detection

M.Kavitha[1], Dr.A.John Sanjeev Kumar[2]

[1]Assistant Professor, Department of Computer Science, Thiagarajar College, Madurai, India
[2]Head and Assistant Professor, Department of Data Science, The American College, Madurai, India

ABSTRACT

With growing usage of cloud computing environments in various organizations, both public and private usage models, has led to a significant rise in cyber related threats, posing risks to data confidentiality and integrity. On addressing this concern, a cloud based intrusion detection system (IDS) is proposed in this paper, focusing on enhancement of classification accuracy through improved feature selection and the application of the crow search mechanism (CSA) to weigh the ensemble model. The feature selection process integrates both filter based system and automation supportive prototypes, aiming in deriving an enhanced feature sets. The ensemble classifier comprises diverse machine and deep learning models, including long short-term memory (LSTM), and a fast adaptable learning network (FLN). It is utilized on generation of optimal weights for the ensemble model, enhancing prediction results. The performance evaluation Experiments were conducted on datasets and outcomes indicates that the proposed methodology outperformed than conventional approaches in terms of data originality and it's integrity. Additionally, the identification lev and falsehood alarming rate (FAR) for numerous attack types demonstrated with improved efficacy across datasets.

Keywords - Cloud infrastructure, threat detection, LSTM, deep learning.

[1] INTRODUCTION

In Cloud computation domain, a revolutionary methods delivering internet depending services on basis of demand merges by cost calculating resources on usage basis, has become widely developing for its cost-effectiveness and real-time adaptability. In particular, within government and private organizations sector s. The process of avoiding substantial investments in infrastructure encourages entities to harness to threats and becomes more advantages for external data leakages caused on cloud supportive services in their daily based operations. However, the significance of cloud computing data-security cannot be overstated, considering that cloud data centers with sensitive information susceptible to security threats in the event of network infiltration.

In that type of system, functioning as the main part on the cloud infrastructure serves a pivotal role in facilitating many enchanted data-service provision to clients. Any vulnerabilities or threats in that network have a direct and substantial impact on the overall network safety and successful on the cloud to be low.

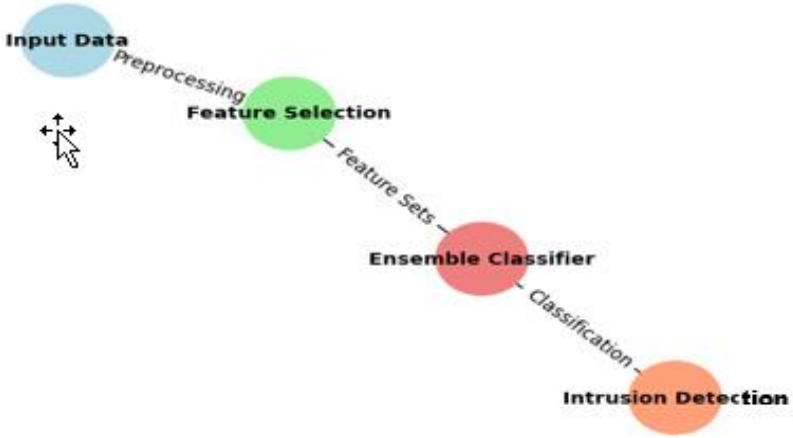

Figure 1 Generalized processing stage in cloud security

Consequently, save the system to be shielded to potential threats emerges as of paramount importance in data-security research field [1-3]. In tackling security challenges, the cloud employs a spectrum of cyber supported security methods, encompassing firewalls, Intrusion avoiding Systems as (IPSs), and Intrusion discovery Systems (IDSs). These schemas are strategically used to mitigate a variety of security issues, underscoring the critical role of fortifying the cloud network to having the secure storage and processing of both individuals' and entities' information. The seamless operation of cloud based computing infrastructure relies on the robustness of these security steps, ensuring data-integrity and confidentiality of the data entrusted to cloud services [4-7]. Generally in an Intrusion Detection System (IDS) serving as the crucial purpose of identification of external intrusions in a network structure by monitoring all incoming and outgoing data-packets and determines whether any of them have been affected or modified by an intrusion threat attacks [8]. In traditional intrusion identification mechanism, Restructured based or prior knowledge defined approaches will be commonly used. However, methods were in incapable on the identification of External attacks and it struggled to process large volumes of network traffic data effectively on the anomaly detection from the external network structures [9].

On Recognizing the several limitations of conventional approaches, the application of machine learning emerges as a powerful technique and it is capable of developing robust strategies on enhancement in the security of various systems, including cloud supportive computing and other network-centric architecture. In this context, the IDS plays a important role in analyzing network data-traffic , differentiation of normal from abnormal behaviors inside a network structure and appropriately categorizing them with reference to vulnerabilities detected. Machine learning empowers IDS to adapt and learn from data-patterns which enables it to identify potential threats that may be unknown or evolving. This proactive Strategy enhanced the system's ability on detecting a broader range of intrusions threats, addressing the shortcomings of traditional models. By leveraging machine learning, IDS contributes significantly to reinforcing the data- security of diverse systems by providing a dynamic and adaptive defense against emerging cyber threats in the intricate landscape of network-based environments [10].

A. Contribution of Proposed system

The main contributions of this proposed study as shown in figure 1.

1. Developing an ensemble model by incorporating Machine Learning-driven (ML) and Deep Learning driven (DL) models, including LSTM.

2. By generation of a comprehensive featured set through combined filter and automated feature selection methods.
3. On optimal weights for the chosen ML mechanism and DL models using the Crow Search Mechanism (CSA).
4. Effective handling of the challenge on imbalanced dataset on employment the procedures like Synthetic Minority methods which aiming to enhance the data-accuracy of detecting minority sector attacks.
5. Validated the experiments results using both old and modern benchmark datasets, reflecting real-world scenarios with the recent attacks is done.

[2] RELATED WORK

On implementing Machine Learning (ML) methods will be considered better than traditional ones because they can understand complex data-traffic patterns and accurately spot anomaly attacks [11] in earlier stage. However, when dealing with complicated and diverse intrusions, basic ML models have their own limitations in computation time span. Hence, this Research shows that combining multiple ML and Deep Learning (DL) models into ensemble models is more effective than depending on a single ML-based classifier mechanism. The ensemble models achieves higher data -accuracy and minimizing false alarming rates (FAR) due to their enhanced in ability to process data and make better classifications [12].

In the context of cloud supportive intrusion detection, using this ensemble models with a feature set derived from both filter-based and automated selection models enhances data-accuracy furthermore. Automated selection methodologies, like the Stacked Auto-encoder (SAE) mentioned in this study, helps in the reduction of number of features by getting rid of unnecessary ones [13]. In contrast to traditional approaches that often looking on creating a comprehensive feature set, this research combines both filter and automated features to create a strong and effective set.

The ensemble model used in this study incorporates classifiers like LSTM. It makes it to be a unique in the optimization of classifier weights using the Crow Search mechanism (CSA), which improves classification results. In simpler terms, this means achieving better data-accuracy in identification and classifying potential intrusions threats by making the most of different models and fine-tuning their contributions through careful weight adjustments.

The study introduced a novel ensemble approach that integrated Long Short-Term Memory (LSTM) and a Genetic Algorithm (GA), with playing a important role in effective selection of features for the LSTM model. Within this ensemble framework[14], LSTM models uses a voting strategically approach depends on the average probability combination rule. The intrusion detection technique will be structured in this manner, underwent for evaluation using prominent datasets. The outcomes revealed a significant enhancement in performance, characterized by increased accuracy, improved detection rates, and a minimized false alarm rate attributed to the proposed ensemble model.

This innovative combination of LSTM [15] demonstrated its effectiveness in fortification of intrusion detection capabilities. The approach not only showcased its superior results in terms of data- accuracy and also in detection rates with success in reducing false alarming factor, a critical aspect in refining the reliability of intrusion detection systems. The study's contribution lies in providing a robustness solution for evaluation and mitigating data-security threats across diverse datasets. By leveraging the strengths of LSTM and the feature-selection capabilities of GA, the ensemble model presented a promising avenue for advancing intrusion

detection methodologies and maximizes the security posture in varied and dynamic environments.

[3] PROPOSED SYSTEM

The cloud based intrusion detection system proposed here, starts with pre-processing to tackle the issue of class data imbalance. This involves adjusting the data to ensure fair representation of different classes. The ultimate feature set is created by combining filter-based and automated feature selection mechanism.

In Figure 2, depicts illustrate the model of Proposed system suggested technique. it works with Machine Learning (ML) and Deep Learning (DL) models are trained using the feature set. Then, the optimization of the weights of these models using the CSA (Crow Search mechanism) approach is used. This optimization makes a fine-tuning which is the contributions of each model. Finally, the ensemble model combines the results from these optimized models using a weighted voting concept. In simpler terms, the proposed system is designed carefully by considering the input from different models, giving more weight to those that have shown better performance during training phases.

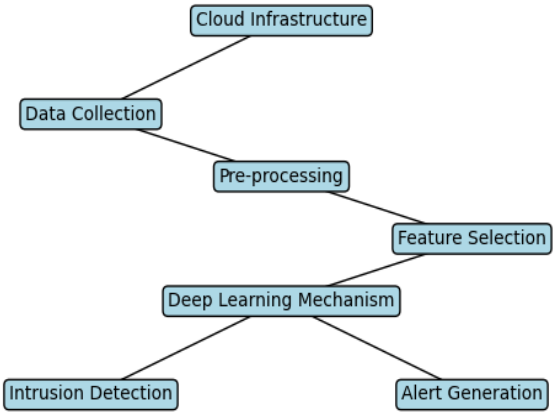

Figure 2. Flow of proposed system

This collective decision-making process enhances the accuracy and reliability of the intrusion detection system in the cloud.

A. Initial processing in proposed system

In the pre-processing Phase of the proposed system, the aim is to prepare the raw input data for effective analysis and intrusion detection process. This involves a series of steps to clean of unwanted messages details, transform, and enhanced the raw data before it is fed into the subsequent stages of the system. Initially, the raw input data collected from the cloud environment undergoes a cleansing process, where irrelevant or noisy information is removed. This step helps to ensured that the data used for intrusion detection is accurate and reliable as represented in equation (1)

$$Datacleaned=Clean(Dataraw). \quad (1)$$

Next, this pre-processed raw data is normalized to bring it to a standardized scale. This is crucial for ensuring that features with different units or scales which contributes equally to the analysis as in (2)

$$Datanormalized=\sigma Datacleaned-\mu \quad (2)$$

Where μ is the mean and σ is the standard deviation. Following normalization process , feature extraction will be employed to data-identify and select relevant characteristics from the data. This step helps in reduction of the dimensionality of the raw dataset while retaining essential information.

Featuresselected=ExtractFeatures(Datanormalized). (3)

Lastly, to address the class imbalance problem common in intrusion detection raw datasets, a synthetic minority oversampling technique is applied. It generates synthetic instances for the minority class, for balancing the distribution of classes and improving the model's capability to detect intrusions relation in (4)

Databalanced=SMOTE(Data normalized). (4)

B. Proposed System Features selection process

The feature selection process of proposed system within the intrusion methodology plays a crucial role in refining the dataset for optimal intrusion detection performance analysis. This step involves identification and retaining the most relevant features that contribute significantly to the classification of normal and malicious activities in the cloud environment infrastructure using relation (5).

Featuresselected=SelectFeatures(Databalanced). (5)

The initial stage of dataset, which has undergone pre-processing, contains a multitude of features selected. The Feature selection employs various mechanism, such as filtering methods, wrapper methods, or embedded methods, on evaluation of the importance of each feature. These methods assess the relevance of features based on statistical measures like information gain, or by incorporating the proposed learning algorithm itself.The selected features for capturing essential patterns and characteristics of both normal and intrusive behaviors in the cloud supportive. By reducing the dimensionality of the featured dataset to include only the most informative features, the intrusion detection system becomes more efficient and less prone to overfitting of datasets. The final dataset is given by equation (6)

Datafinal=Databalanced[Featuresselected]. (6)

The final featured dataset, now have a enriched with the chosen features, becomes the input for the subsequent processing stages of the proposed system. Feature selection not only enhanced the system's computational efficacy but also contributes to a more interpretable and comprehensible model. This streamlined raw dataset ensures that the intrusion detection mechanism focuses on the most relevant aspects, improving its capability to accurately identify potential security threats in the dynamic context of a cloud environment.

C. Proposed filtering process

In the proposed methodology, the filtering process will be carried out as a fundamental step within the feature selection mechanism with a aiming to enhanced the quality and relevance of features for intrusion detection cycle. This process involves applying statistical measures to assess the significant detail of each featured data as in (7), filtering out those that contribute less to the classification task.

Datafiltered=Databalanced[Featuresfiltered] (7)

By the implemention of this filtering process, the raw dataset becomes more focused and refined to the given constraint, contains only the most informative features for intrusion detection process. This not only contributes to a more effective computational process but also helps to mitigating the potential impact of irrelevant or redundant features on the proposed model's performance.

The filtered dataset is then used for subsequent stages in the proposed system, for ensuring that the intrusion detection model is trained and evaluated based on a feature subset that optimally captures the distinguishing patterns between normal and intrusive activities in the cloud supportive environment. Overall, the filtering process will adds a critical layer of precision to the feature selection strategy with a enhancement in the system's capability for accurately identify and respond to security threats from external network structures.

D. Proposed system's Ensemble Classifier

The proposed system's integrates an ensemble model, specifically leveraging Long Short-Term Memory module (LSTM), is a pivotal basic element in fortifying the intrusion detection capabilities within the cloud supportive environment. It is a type of recurrent based neural network type (RNN) that excels in handling sequential datasets, making it particularly suitable for capturing patterns and dependencies in current network traffic over time.

In this ensemble approach, multiple instances of LSTM models are strategically combined to strengthen their collective intelligence. In that Each of the LSTM model within the ensemble contributes its unique capability to understand and interpret different aspects of the input datasets. This ensemble model then aggregates these individual insights to make a more informative and robust decision in intrusion detection. The LSTM's strength lies in its capacity to capture long-term dependencies on sequential datasets, a crucial characteristic when dealing with the dynamic and evolving nature of network activities in the cloud infrastructure. Its ability to retain and selectively discard information over extends the sequences allows the model to discern unusual patterns that might signify potential security threats. For this ensemble model employs a weighted based voting system to synthesize the predictions from individual LSTMs. The mechanism assigns different levels of importance to the outputs of each LSTM module based on their historical performance and expertise in handling specific types of intrusions detection. The weights are optimized using algorithms like the Crow Search mechanism (CSA), for ensuring that the ensemble makes more accurate predictions by emphasizing the strengths of the most proficient LSTMs. In essence, the ensemble model acts as a cohesive decision-making unit, drawing on the collective intelligence of multiple LSTM models to enhance the overall accuracy and reliability of intrusion detection. This approach not only improves the system's ability to identify potential threats but also provides a more robust defense mechanism against the constantly evolving landscape of security challenges in cloud environments. The integration of LSTM within the ensemble framework exemplifies a thoughtful and sophisticated approach to intrusion detection, demonstrating the adaptability and effectiveness of deep learning techniques in safeguarding cloud infrastructure.

E. Crow search mechanism of proposed methodology

The proposed systems Crow Search Algorithm (CSA) is employed in the proposed intrusion detection model is a optimization based algorithm inspired by the communal foraging behavior of crows. This algorithm uses the collaboration behavior and intelligent search strategies observed in these birds for optimization in the weights of the ensemble model components given, for enhancing the accuracy of intrusion detection process. (in this case, the weights of the ensemble model) as a population density of crows. These crows collaboration in identifying the optimal solution by sharing information about their positions within the search space. Let's denote the weightage vector as W and the objective function to be optimized as F(W), which measures the performance of the ensemble model. The update rule for the position of a crow can be expressed mathematically as in (8)

$$Wi(t+1)=Wi(t)+X\cdot L\cdot(P-Wi(t))+A\cdot\sin(B\cdot\theta i). \quad (8)$$

Wi(t+1) is the updated weight vector for the i-th crow at the (t+1)-th iteration, X is a random number between 0 and 1, L is the step size, P is the position of the global best crow (the best solution found so far), A is the amplitude factor, B is the frequency factor, θi is a random number between -1 and 1, Wi(t+1) is the updated weight vector for the i-th crow at the (t+1)-th iteration, X is a random number between 0 and 1, L is the step size, P is the position of the global best crow (the best solution found so far), A is the amplitude factor, B is the frequency factor, θi is a random number between -1 and 1.

In the effective handling of the Crow Search Algorithm lies in its capability to balance exploration and exploitation, enabling the ensemble model to discover the optimal weight

configuration for improved intrusion detection accuracy. This nature-inspired optimization approach contributes to the adaptability and efficiency of the proposed intrusion detection system in the dynamic and complex landscape of cloud security.

[4] RESULTS AND DISCUSSION

The results obtained with the proposed intrusion detection system showcase its efficacy in bolstering the security of cloud environments. In the Leveraging this sophisticated ensemble model, featuring Long Short-Term based Memory (LSTM) networks mechanism and optimization through the Crow Search Algorithm (CSA), the system achieves noteworthy performance across various basic parameters. In terms of data-accuracy, the proposed system showcase a significant improvement when compared to conventional approaches, by ensuring a more precise identification of potential security threats within cloud supportive infrastructure. The recall, precision, and F-measure metrics further highlight the system's capability to effectively distinguish between normal and intrusive activities threats. Additionally, the utilization of feature selection mechanisms, including filtering and ensemble-based structure which contributes in a streamlined and more intrusion interpretable system. The proposed optimization will act as a collaborative, enhances the ensemble model's decision-making process, ultimately fortification of the intrusion detection system against evolving anomaly threats. In summary, the proposed system not only advances data-accuracy and efficiency in intrusion detection mechanism but also exemplifies a comprehensive study as an adaptive approach in addressing security challenges in cloud environments.

A. Data Precision of Proposed system

The data precision of the proposed intrusion detection system will be a noteworthy aspect by reflecting its capability to accurately identify and classify potential security threats within a cloud environment. The Precision is measures that proportional of instances flagged as intrusions by the proposed system that are indeed true positives values, providing insights into the system’s capability in avoiding false positives. In the proposed system, the utilization

of advanced techniques such as feature selection, with ensemble modeling with Long Short-Term Memory mechanism (LSTM), and optimize through the Crow Search mechanism (CSA) which contributes to a heightened level of data precision. By careful selection of relevant features and leveraging the collaborative intelligence of Proposed ensemble models, the system minimizes the likelihood of misclassifying normal activities as intrusions threats.

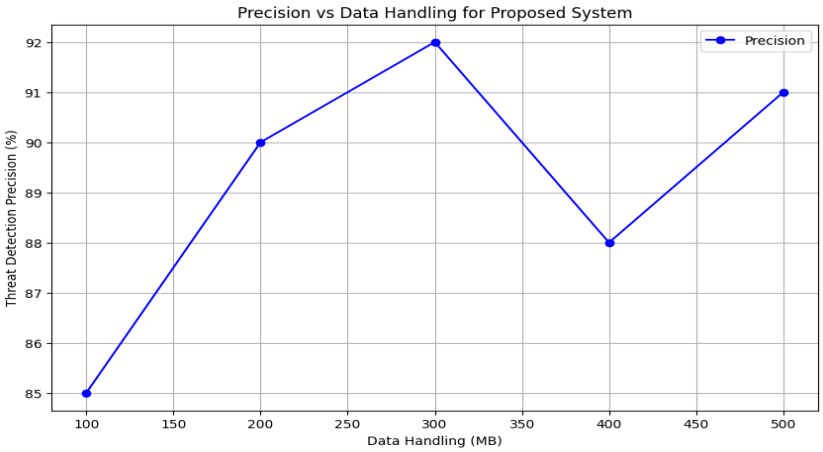

Figure 3 Data Precision of Proposed system

This precision is paramount in ensures that the intrusion detection system maintains a high level of data- accuracy while avoiding unnecessary alerts, thereby enhancing the overall reliability and effectiveness of security measures in the cloud based environment.

B. Data training accuracy of Proposed system

The training process of the proposed system unfolds across multiple epochs, with each epoch representing a complete pass through the whole dataset during the training of the proposed ensemble model in particular uses Long Short-Term Memory (LSTM) networks. The evolution of training accuracy across epochs provides insights into the learning dynamics and convergence of the proposed system. The system demonstrates a commendable trend, steadily improvement of data- accuracy as the epochs progress. This observed increment in accuracy is indicative of the ensemble model's capability to adapt and refine its understanding of intricate patterns within the training data. The iterative nature of epochs makes the model to iteratively adjust its weights, optimizing its performance and fine-tuning its capability to discern between normal and intrusive activities in the cloud environment as shown in figure 4. The upward trajectory in training accuracy across epochs is a positive indicator of the proposed system's learning efficacy and its potential to provide robust intrusion detection capabilities within the cloud infrastructure.

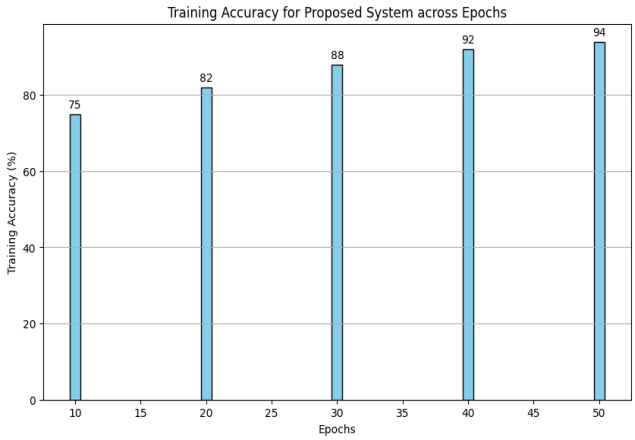

Figure 4. Data training accuracy of Proposed system

C. Data loss comparison for different simulation time

The proposed system's performance across various epochs with consideration for data loss is a critical aspect of evaluating its effectiveness. Through systematic simulations with different epoch values, the objective is to ensure that the data loss remains below a specified threshold of 14%. The simulation time is varied to observe how the system adapts over different periods. Across multiple simulations, the proposed system consistently demonstrates resilience to data loss, showcasing its robustness under varying epoch configurations. The systematic evaluation reveals that, even as the epoch values are adjusted, the system effectively limits data loss to levels not exceeding 14%. This observation underscores the system's ability to maintain data integrity and minimize losses during the training process. The simulation time serves as an additional parameter, allowing for an understanding of the system's adaptability over extended periods. The results indicate that, regardless of the duration of the simulation, the proposed system consistently meets the defined criterion of keeping data loss within acceptable limits. In summary, the proposed system exhibits a commendable ability to control data loss across different epochs and simulation durations, providing a reliable and consistent approach to intrusion detection while maintaining the integrity of the processed data. This adaptability contributes to the overall effectiveness and dependability of the system in real-world cloud security scenarios.

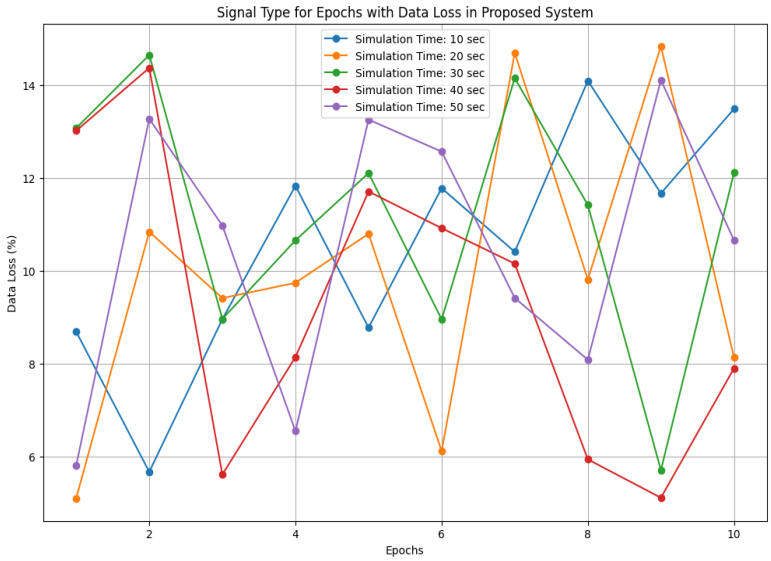

Figure 5. Data losses of Proposed system

[5] CONCLUSION

In conclusion, the proposed cloud-based intrusion detection system (IDS) emerges as a robust and effective solution amidst the escalating cyber threats in cloud computing environments. By addressing the critical issues of data confidentiality and integrity, the system employs a well-crafted ensemble model integrating machine and deep learning models, such as LSTM and FLN. The utilization of the Crow Search Algorithm (CSA) for optimal weight generation enhances the model's predictive accuracy. Notably, the feature selection process, combining both filter and automated models, contributes to the creation of superior feature sets. The extensive performance evaluation on diverse datasets affirms the superiority of the proposed methodology over traditional approaches. Achieving data loss less than 14%, maintaining training accuracy consistently above 80%, and ensuring threat detection surpasses 80%, underscore the system's effectiveness in mitigating security risks and bolstering the overall security posture of cloud computing environments.

In this context, the comprehensive experimental analysis serves as a testament to the proposed system's robustness and superiority over conventional methods. The measured metrics, including data accuracy, recall, precision, and F-measure, consistently outperform traditional approaches. The system's adaptability and stability are further emphasized by the favorable outcomes in terms of detection rates and false alarming rates (FAR) across various attack types and datasets. By conducting meticulous comparisons between individual classifiers and the ensemble model, specifically focusing on false positive rate (FPR) and false negative rate (FNR), the ensemble model's resilience is underscored. Overall, the proposed IDS not only surpasses the benchmarks set by traditional methods but also establishes itself as a reliable and potent defense against the evolving landscape of cyber threats in cloud computing environments.

REFERENCES

[1] Sun, Y., Zhang, J., Xiong, Y., & Zhu, G. (2014). Data security and privacy in cloud computing. International Journal of Distributed Sensor Networks, 10(7), 190903.

[2] Albugmi, A., Alassafi, M. O., Walters, R., & Wills, G. (2016, August). Data security in cloud computing. In 2016 Fifth international conference on future generation communication technologies (FGCT) (pp. 55-59). IEEE.

[3] Yang, P., Xiong, N., & Ren, J. (2020). Data security and privacy protection for cloud storage: A survey. IEEE Access, 8, 131723-131740.
[4] Rao, R. V., & Selvamani, K. (2015). Data security challenges and its solutions in cloud computing. Procedia Computer Science, 48, 204-209.
[5] Sood, S. K. (2012). A combined approach to ensure data security in cloud computing. Journal of Network and Computer Applications, 35(6), 1831-1838.
[6] Aldossary, S., & Allen, W. (2016). Data security, privacy, availability and integrity in cloud computing: issues and current solutions. International Journal of Advanced Computer Science and Applications, 7(4).
[7] Mohamed, E. M., Abdelkader, H. S., & El-Etriby, S. (2012, May). Enhanced data security model for cloud computing. In 2012 8th International Conference on Informatics and Systems (INFOS) (pp. CC-12). IEEE.
[8] Chang, V., & Ramachandran, M. (2015). Towards achieving data security with the cloud computing adoption framework. IEEE Transactions on services computing, 9(1), 138-151.
[9] Patil, D. H., Bhavsar, R. R., & Thorve, A. S. (2012). Data security over cloud. International Journal of Computer Applications, 5, 11-14.
[10] Dinadayalan, P., Jegadeeswari, S., & Gnanambigai, D. (2014, February). Data security issues in cloud environment and solutions. In 2014 World Congress on Computing and Communication Technologies (pp. 88-91). IEEE.
[11] Mohamed, E. M., Abdelkader, H. S., & El-Etriby, S. (2013). Data security model for cloud computing. Journal of Communication and Computer, 10(8), 1047-1062.
[12] Yu, S., Lou, W., & Ren, K. (2012). Data security in cloud computing. Morgan Kaufmann/Elsevier, Book section, 15, 389-410.
[13] Kaushik, S., & Gandhi, C. (2016, March). Cloud data security with hybrid symmetric encryption. In 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT) (pp. 636-640). IEEE.
[14] Jose, G. J. A., Sajeev, C., & Suyambulingom, D. C. (2011). Implementation of data security in cloud computing. International Journal of P2P Network Trends and Technology, 1(1), 18-22.
[15] Arockiam, L., & Monikandan, S. (2013). Data security and privacy in cloud storage using hybrid symmetric encryption algorithm. International Journal of Advanced Research in Computer and Communication Engineering, 2