

QUANTIFYING CYBER RISK: PREDICTIVE MODELS FOR IDENTIFYING AND PREVENTING HACKING BREACHES

Dr Kondragunta Rama Krishnaiah¹, Dr Muvva Venkateswara Rao²

¹Professor, Dept. of CSE, R K College of Engineering, Vijayawada-521456, A.P, India.

²Professor, Dept. of CSE, NRI Institute of Technology, Vijaywada- 521212, India.

ABSTRACT

With the escalating complexity of cyber threats, analyzing incident data sets has emerged as a crucial avenue for enhancing our comprehension of the evolving threat landscape. This research delves into a relatively new and imperative field, focusing on a 12-year span from 2005 to 2017, encapsulating cyber hacking activities. A comprehensive statistical analysis of breach incident data sets forms the core of our investigation. We propose specific stochastic process models tailored to fit the inter-arrival times and breach sizes, crucial dimensions in understanding cyber threat dynamics. Our research demonstrates the efficacy of these models in predicting both inter-arrival times and breach sizes, marking a significant advancement in proactive cyber security measures.

Qualitative and quantitative trend analyses are conducted on the dataset, offering nuanced insights into the evolving nature of cyber threats over the studied period. The amalgamation of statistical rigor, stochastic modeling, and trend analyses contributes to a holistic understanding of cyber threat evolution, paving the way for enhanced predictive capabilities and proactive cyber security strategies. This research underscores the urgency for further exploration in this nascent field, recognizing the plethora of opportunities to refine our understanding of cyber threats and fortify digital defenses.

Keywords: Cyber Threats, Cyber Risk Analysis, Breach Prediction, Trend analysis.

1. INTRODUCTION

CONCEPTUAL BACKGROUND:

Data breaches stand out as among the most devastating cyber incidents, persisting as a significant challenge despite advancements in technological defenses. While cybersecurity measures aim to fortify systems against attacks, the prevalence of data breaches remains a pressing concern. Recent research endeavors have turned towards modeling data breach incidents, recognizing the need to understand and address the human errors associated with negligent breaches. Motivated by unexplored questions in the field, this study seeks to uncover trends in data breaches, particularly focusing on whether incidents caused by cyber-attacks are increasing, decreasing, or stabilizing. In contrast to existing literature, our investigation reveals that modeling hacking breach incident inter-arrival times and breach sizes demands a departure from traditional distributions, as these factors exhibit notable autocorrelations.

This study introduces specific stochastic process models tailored to capture the intricacies of both hacking breach incident inter-arrival times and breach sizes. Demonstrating the predictive capabilities of these models, we illuminate their potential to anticipate future incidents. In our pursuit of a deeper understanding of the evolution of hacking breach incidents, we employ a dual approach, conducting both qualitative and quantitative trend analyses on the dataset. The insights drawn from our analyses contribute to the cybersecurity discourse. Contrary to some existing findings, our study suggests that the frequency of cyber hacks is on the rise, emphasizing the need for heightened vigilance. However, our research also reveals that the magnitude of damage caused by these incidents may not necessarily be increasing at a proportional rate. These nuanced findings pave the way for more targeted and effective cybersecurity strategies, acknowledging the evolving nature of cyber threats and the imperative to stay ahead of malicious actors.

2. LITERATURE REVIEW

Analyzing cyber incident datasets is a crucial approach for enhancing our comprehension of the evolving threat landscape. This research area, while relatively new, holds great promise, with numerous avenues for further exploration. In particular, breach sizes in cybersecurity incidents should be modeled using stochastic processes. Notably, negligent breaches, often stemming from human errors, present a significant aspect of the cybersecurity landscape, distinct from intentional cyber-attacks. As we delve into this emerging field, it becomes evident that stochastic processes provide a more fitting framework for modeling breach sizes. Unlike traditional distribution-based models, stochastic processes better capture the dynamic and evolving nature of cybersecurity incidents, especially considering the varying impact of negligent breaches and intentional cyber-attacks. Human errors leading to negligent breaches introduce a unique dimension to the cybersecurity challenge. Understanding the dynamics of these incidents is crucial for developing effective mitigation and prevention strategies. While technological solutions are pivotal in

defending against intentional cyber-attacks, addressing human errors requires a nuanced approach that considers the behavioral and organizational aspects contributing to these incidents.

In the context of existing systems, the integration of stochastic processes for modeling breach sizes offers an opportunity to refine risk assessments and incident response strategies. By acknowledging the probabilistic and evolving nature of cybersecurity incidents, organizations can better prepare for and mitigate the impact of breaches. Moving forward, there is a need for further research to explore the intricacies of modeling breach sizes using stochastic processes. This includes refining models based on real-world incident data, identifying key factors influencing breach sizes, and developing predictive capabilities to enhance cybersecurity resilience. As the cybersecurity landscape continues to evolve, embracing stochastic modeling becomes essential for staying ahead of emerging threats and ensuring the robustness of defense mechanisms within existing systems.

3. PROPOSED APPROACH

Investigate the Cybercrime Underground Economy Emphasize the closed community nature of the cybercrime underground. Collect Data Relevant to the Cybercrime Underground. Consider what data is needed and where it can be obtained. Collect data from the cybercrime underground community. Obtain malware-related data from a leading global cyber security research firm. Emphasize the importance of Remote Access Trojans (RATs). Integrate RAT elements into all steps (Steps 1-4) of the framework. Model hacking breach incident interarrival times and breach sizes. Utilize ARMA-GARCH models to describe the evolution of hacking breaches. Stress the use of stochastic processes over distributions. Consider dependence for accurate prediction results in inter-arrival times and breach sizes. A comprehensive approach covering all phases of data analysis. Emphasis on RATs throughout the process. Integration of data collection, modeling, and application development. Develop an application based on the analytical methods. Implement the application to facilitate the investigation of the cybercrime underground. Ensure the application incorporates RAT-based definitions and insights gained from the data analysis.



FIGURE 1: Identifying And Preventing Hacking Breaches

3.1 ALGORITHM:

SUPPORT VECTOR MACHINE (S.V.M): Support Vector Machine (SVM) is a supervised machine learning algorithm.

Which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the below snapshot). Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line). More formally, a support vector machine constructs a hyper plane or set of hyper planes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training- data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. Whereas the original problem may be stated in a finite dimensional space, it often

happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space.

3.2 METHODS:

First, we show that both the hacking breach incident interarrival times (reflecting incident frequency) and breach sizes should be modeled by stochastic processes, rather than by distributions.

- To describe the evolution of hacking breaches we use ARMA-GARCH model methods.
- These stochastic process models can predict the inter-arrival times and the breach sizes.
- When predicting inter-arrival times and breach sizes, it is necessary to consider the dependence; otherwise, the prediction results are not accurate.

4. RESULTS AND DISCUSSION

In (left panel) there is a clear trend of 1367 daily observations prices of the S.M.R 20 in Malaysia from 4th January 2010 to 4th August 2015. When daily prices converted to log returns, the plot in (right panel) illustrates there are large negative values especially on March and October 2011.



FIGURE 2: Daily Prices For The Smr20

4.1 TREND ANALYSIS:

The presentation of both qualitative and quantitative trend analyses on hacking breach incidents based on the models presented earlier. The approach involves decomposing the data into two parts: the trend part and the random (or noise) part. By combining both qualitative and quantitative analyses, this approach aims to provide a comprehensive understanding of the trends in hacking breach incidents, considering both the statistical patterns and the context in which they occur.

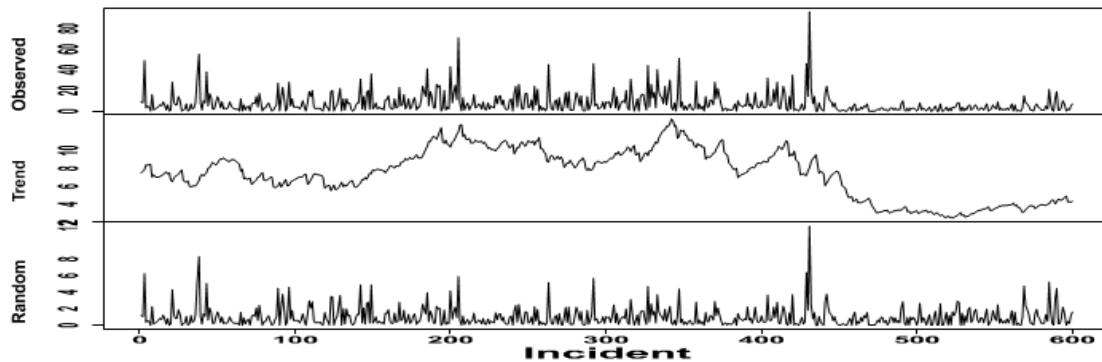


Figure 3: both qualitative and quantitative trend analyses on the hacking breach

It appears there might be a discrepancy between the finding presented and the previous conclusion that was based on a super dataset involving different incident types (negligent breaches and malicious breaching). The present study, however, seems to concentrate specifically on hacking breach incidents, which is a subset or a proper sub-type within the broader category of malicious breaches. By explicitly addressing the scope difference and contextualizing the findings, you can ensure a clear and accurate representation of the insights derived from the specific focus on hacking breach incidents in the present study.

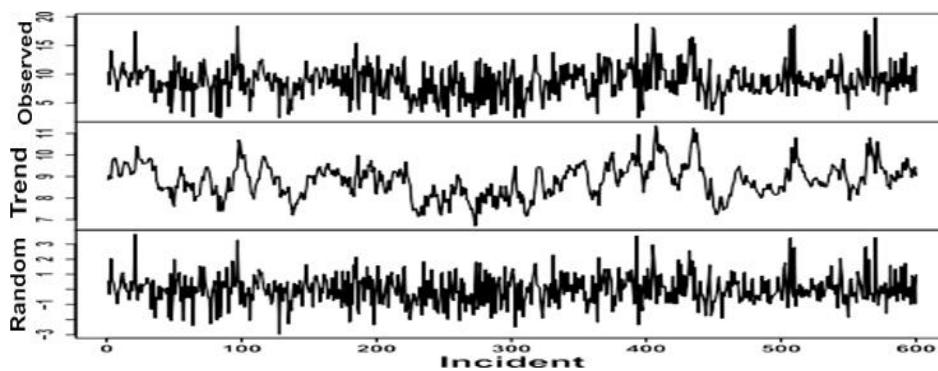


Figure 4: Using the ARMA-GARCH model to decompose the log-TRANSFORMED BREACH SIZES INTO A TREND part and a random part

5. CONCLUSION

We analyzed a hacking breach dataset from the points of view of the incidents inter-arrival time and the breach size, and showed that they both should be modeled by stochastic processes rather than distributions. The statistical models developed in this paper show satisfactory fitting and prediction accuracies. In particular, we propose using a copula-based approach to predict the joint probability that an incident with a certain magnitude of breach size will occur during a future period of time. Statistical tests show that the methodologies proposed in this paper are better than those which are presented in the literature, because the latter ignored both the temporal correlations and the dependence between the incidents inter-arrival times and the breach sizes. We conducted qualitative and quantitative analyses to draw further insights. We drew a set of cyber security insights, including that the threat of cyber hacking breach incidents is indeed getting worse in terms of their frequency, but not the magnitude of their damage. The methodology presented in this paper can be adopted or adapted to analyze datasets of a similar nature.

5.1 FUTURESCOPE:

Real-time intelligence: The longer it takes to identify a hack, the more costly its consequences. With just 60 seconds' notification of a compromise, resulting costs could be reduced by 40%.

Cyber-insurance: Insurers typically limit their capacity to between \$5 million and \$100 million per client. As of October 2016, only 29% of US business had purchased cyber-insurance. However, the overall cyber-insurance market is estimated to be \$20 billion by 2025, up from \$3.25 billion today.

Bug bounty programs: Organizations pay outsiders ("friendly hackers") to notify them of security flaws. Companies ranging from Google and Dropbox to AT&T and LinkedIn have already adopted this practice.

6. REFERENCES

- [1] Kommineni, K. K. ., & Prasad, A. . (2023). A Review on Privacy and Security Improvement Mechanisms in MANETs. International Journal of Intelligent Systems and Applications in Engineering, 12(2), 90–99. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/4224>
- [2] Vellela, S.S., Balamanigandan, R. Optimized clustering routing framework to maintain the optimal energy status in the wsn mobile cloud environment. *Multimed Tools Appl* (2023). <https://doi.org/10.1007/s11042-023-15926-5>
- [3] Vellela, S. S., Reddy, B. V., Chaitanya, K. K., & Rao, M. V. (2023, January). An Integrated Approach to Improve E-Healthcare System using Dynamic Cloud Computing Platform. In 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 776-782). IEEE.
- [4] K. N. Rao, B. R. Gandhi, M. V. Rao, S. Javvadi, S. S. Vellela and S. Khader Basha, "Prediction and Classification of Alzheimer's Disease using Machine Learning Techniques in 3D MR Images," 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2023, pp. 85-90, doi: 10.1109/ICSCSS57650.2023.10169550.
- [5] VenkateswaraRao, M., Vellela, S., Reddy, V., Vullam, N., Sk, K. B., & Roja, D. (2023, March). Credit Investigation and Comprehensive Risk Management System based Big Data Analytics in Commercial Banking. In 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS) (Vol. 1, pp. 2387-2391). IEEE [6]
- [6] S Phani Praveen, RajeswariNakka, AnuradhaChokka, VenkataNagarajuThatha, SaiSrinivasVellela and UddagiriSirisha, "A Novel Classification Approach for Grape Leaf Disease Detection Based on Different Attention Deep Learning Techniques" International Journal of Advanced Computer Science and Applications(IJACSA), 14(6), 2023. <http://dx.doi.org/10.14569/IJACSA.2023.01406128>

[7] Vellela, S. S., & Balamanigandan, R. (2022, December). Design of Hybrid Authentication Protocol for High Secure Applications in Cloud Environments. In 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS) (pp. 408-414). IEEE.

[8] Vullam, N., Vellela, S. S., Reddy, V., Rao, M. V., SK, K. B., & Roja, D. (2023, May). Multi-Agent Personalized Recommendation System in E-Commerce based on User. In 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) (pp. 1194-1199). IEEE.

[9] Vellela, S. S., Balamanigandan, R., & Praveen, S. P. (2022). Strategic Survey on Security and Privacy Methods of Cloud Computing Environment. Journal of Next Generation Technology (ISSN: 2583-021X), 2(1).

[10] Vellela, S. S., & Krishna, A. M. (2020). On Board Artificial Intelligence With Service Aggregation for Edge Computing in Industrial Applications. Journal of Critical Reviews, 7(07), 2020.

[11] Madhuri, A., Jyothi, V. E., Praveen, S. P., Sindhura, S., Srinivas, V. S., & Kumar, D. L. S. (2022). A New Multi-Level Semi-Supervised Learning Approach for Network Intrusion Detection System Based on the 'GOA'. Journal of Interconnection Networks, 2143047.

[12] Madhuri, A., Praveen, S. P., Kumar, D. L. S., Sindhura, S., & Vellela, S. S. (2021). Challenges and issues of data analytics in emerging scenarios for big data, cloud and image mining. Annals of the Romanian Society for Cell Biology, 412-423.

[13] Praveen, S. P., Sarala, P., Kumar, T. K. M., Manuri, S. G., Srinivas, V. S., & Swapna, D. (2022, November). An Adaptive Load Balancing Technique for Multi SDN Controllers. In 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAIS) (pp. 1403-1409). IEEE.

[14] Vellela, S. S., Basha Sk, K., & Yakubreddy, K. (2023). Cloud-hosted concept-hierarchy flex-based infringement checking system. International Advanced Research Journal in Science, Engineering and Technology, 10(3).

[15] Rao, M. V., Vellela, S. S., Sk, K. B., Venkateswara, R. B., & Roja, D. (2023). SYSTEMATIC REVIEW ON SOFTWARE APPLICATION UNDERDISTRIBUTED DENIAL OF SERVICE ATTACKS FOR GROUP WEBSITES. Dogo Rangsang Research Journal UGC Care Group I Journal, 13(3), 2347-7180.

[16] Venkateswara Reddy, B., Vellela, S. S., Sk, K. B., Roja, D., Yakubreddy, K., & Rao, M. V. Conceptual Hierarchies for Efficient Query Results Navigation. International Journal of All Research Education and Scientific Methods (IJARESM), ISSN, 2455-6211.

[17] Sk, K. B., Roja, D., Priya, S. S., Dalavi, L., Vellela, S. S., & Reddy, V. (2023, March). Coronary Heart Disease Prediction and Classification using Hybrid Machine Learning Algorithms. In 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA) (pp. 1-7). IEEE.

[18] Sk, K. B., & Vellela, S. S. (2019). Diamond Search by Using Block Matching Algorithm. DIAMOND SEARCH BY USING BLOCK MATCHING ALGORITHM. International Journal of Emerging Technologies and Innovative Research (www. jetir. org), ISSN, 2349-5162.

[19] Yakubreddy, K., Vellela, S. S., Sk, K. B., Reddy, V., & Roja, D. (2023). Grape CS-ML Database-Informed Methods for Contemporary Vineyard Management. International Research Journal of Modernization in Engineering Technology and Science, 5(03).

[20] Vellela, Sai Srinivas and Chaganti, Aswini and Gadde, Srimadhuri and Bachina, Padmapriya and Karre, Rohiwalter, A Novel Approach for Detecting Automated Spammers in Twitter (June 24, 2023). Mukt Shabd Journal Volume XI, Issue VI, JUNE/2022 ISSN NO : 2347-3150, pp. 49-53 , Available at SSRN: <https://ssrn.com/abstract=4490635>

[21] Vellela, Sai Srinivas and Pushpalatha, D and Sarathkumar, G and Kavitha, C.H. and Harshithkumar, D, ADVANCED INTELLIGENCE HEALTH INSURANCE COST PREDICTION USING RANDOM FOREST (March 1, 2023). ZKG International, Volume VIII Issue I MARCH 2023, Available at SSRN: <https://ssrn.com/abstract=4473700>

[22] Dalavai, L., Javvadi, S., Sk, K. B., Vellela, S. S., & Vullam, N. (2023). Computerised Image Processing and Pattern Recognition by Using Machine Algorithms.

[23] Vellela, S. S., Basha Sk, K., & Javvadi, S. (2023). MOBILE RFID APPLICATIONS IN LOCATION BASED SERVICES ZONE. MOBILE RFID APPLICATIONS IN LOCATION BASED SERVICES ZONE", International Journal of Emerging Technologies and Innovative Research (www. jetir. org| UGC and issn Approved), ISSN, 2349-5162.

[24] Vellela, Sai Srinivas and Sk, Khader Basha and B, Venkateswara Reddy, Cryonics on the Way to Raising the Dead Using Nanotechnology (June 18, 2023). INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS), Vol. 03, Issue 06, June 2023, pp : 253-257,

[25] Vellela, Sai Srinivas and D, Roja and B, Venkateswara Reddy and Sk, Khader Basha and Rao, Dr M Venkateswara, A New Computer-Based Brain Fingerprinting Technology (June 18, 2023). International Journal

Of Progressive Research In Engineering Management And Science, Vol. 03, Issue 06, June 2023, pp : 247-252 e-ISSN : 2583-1062.,

[26] Gajjala, Buchibabu and Mutyala, Venubabu and Vellela, Sai Srinivas and Pratap, V. Krishna, Efficient Key Generation for Multicast Groups Based on Secret Sharing (June 22, 2011). International Journal of Engineering Research and Applications, Vol. 1, Issue 4, pp.1702-1707, ISSN: 2248-9622

[27] Kiran Kumar Kommineni, Ratna Babu Pilli, K. Tejaswi, P. Venkata Siva, Attention-based Bayesian inferential imagery captioning maker, Materials Today: Proceedings, 2023, ISSN 2214-7853, <https://doi.org/10.1016/j.matpr.2023.05.231>.

[28] Venkateswara Reddy, B., & KhaderBashaSk, R. D. Qos-Aware Video Streaming Based Admission Control And Scheduling For Video Transcoding In Cloud Computing. In International Conference on Automation, Computing and Renewable Systems (ICACRS 2022).

[29] Reddy, N. V. R. S., Chitteti, C., Yesupadam, S., Desanamukula, V. S., Vellela, S. S., & Bommagani, N. J. (2023). Enhanced speckle noise reduction in breast cancer ultrasound imagery using a hybrid deep learning model. Ingénierie des Systèmes d'Information, Vol. 28, No. 4.

[30] Vellela, S. S., & Balamanigandan, R. (2023). An intelligent sleep-aware energy management system for wireless sensor network. Peer-to-Peer Networking and Applications, 16(6), 2714-2731.

[31] Rao, D. M. V., Vellela, S. S., Sk, K. B., & Dalavai, L. (2023). Stematic Review on Software Application Under-distributed Denial of Service Attacks for Group Website. DogoRangsang Research Journal, UGC Care Group I Journal, 13.

[32] Priya, S. S., Vellela, S. S., Reddy, V., Javvadi, S., Sk, K. B., & Roja, D. (2023, June). Design And Implementation of An Integrated IOT Blockchain Framework for Drone Communication. In 2023 3rd International Conference on Intelligent Technologies (CONIT) (pp. 1-5). IEEE.

[33] Vullam, N., Yakubreddy, K., Vellela, S. S., Sk, K. B., Reddy, V., & Priya, S. S. (2023, June). Prediction And Analysis Using A Hybrid Model For Stock Market. In 2023 3rd International Conference on Intelligent Technologies (CONIT) (pp. 1-5). IEEE.

[34] K. K. Kumar, S. G. B. Kumar, S. G. R. Rao and S. S. J. Sydulu, "Safe and high secured ranked keyword search over an outsourced cloud data," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, India, 2017, pp. 20-25, doi: 10.1109/ICICI.2017.8365348.

[35] Sk, K. B., Vellela, S. S., Yakubreddy, K., & Rao, M. V. (2023). Novel and Secure Protocol for Trusted Wireless Ad-hoc Network Creation. Khader Basha Sk, Venkateswara Reddy B, Sai Srinivas Vellela, Kancharakunt Yakub Reddy, M Venkateswara Rao, Novel and Secure Protocol for Trusted Wireless Ad-hoc Network Creation, 10(3).

[36] Vellela, S. S., Sk, K. B., Dalavai, L., Javvadi, S., & Rao, D. M. V. (2023). Introducing the Nano Cars Into the Robotics for the Realistic Movements. International Journal of Progressive Research in Engineering Management and Science (IJPREMS) Vol, 3, 235-240.

[37] Kumar, K. & Babu, B. & Rekha, Y.. (2015). Leverage your data efficiently: Following new trends of information and data security. International Journal of Applied Engineering Research. 10. 33415-33418.

[38] Vellela, S. S., Reddy, V. L., Roja, D., Rao, G. R., Sk, K. B., & Kumar, K. K. (2023, August). A Cloud-Based Smart IoT Platform for Personalized Healthcare Data Gathering and Monitoring System. In 2023 3rd Asian Conference on Innovation in Technology (ASIANCON) (pp. 1-5). IEEE.

[39] Davuluri, S., Kilaru, S., Boppana, V., Rao, M. V., Rao, K. N., & Vellela, S. S. (2023, September). A Novel Approach to Human Iris Recognition And Verification Framework Using Machine Learning Algorithm. In 2023 6th International Conference on Contemporary Computing and Informatics (IC3I) (Vol. 6, pp. 2447-2453). IEEE.

[40] Vellela, S. S., Vuuyuru, L. R., MalleswaraRaoPurimetla, N., Dalavai, L., & Rao, M. V. (2023, September). A Novel Approach to Optimize Prediction Method for Chronic Kidney Disease with the Help of Machine Learning Algorithm. In 2023 6th International Conference on Contemporary Computing and Informatics (IC3I) (Vol. 6, pp. 1677-1681). IEEE.

[41] Vellela, S. S., Roja, D., Sowjanya, C., SK, K. B., Dalavai, L., & Kumar, K. K. (2023, September). Multi-Class Skin Diseases Classification with Color and Texture Features Using Convolution Neural Network. In 2023 6th International Conference on Contemporary Computing and Informatics (IC3I) (Vol. 6, pp. 1682-1687). IEEE.

[42] Vellela, S. S., Sk, K. B., & Reddy, V. An Intelligent Decision Support System for retrieval of patient's information.

[43] Rao, M. V., Seeraman, Y., Mantena, S. V., Gundu, V., Roja, D., & Vatambeti, R. (2023). Brinjal Crop yield prediction using Shuffled shepherd optimization algorithm based ACNN-OBDLSTM model in Smart Agriculture . Journal of Integrated Science and Technology, 12(1), 710. Retrieved from <https://pubs.thesciencein.org/journal/index.php/jist/article/view/a710>