[1]Pokkuluri Kiran Sree

[2]Prasun Chakrabarti

[3]Martin Margala

[4]SSSN Usha Devi N.

Auto encoders with Cellular Automata for Anomaly Detection

Journal of Electrical Systems

Abstract: - This work combines auto encoders with cellular automata (CA) to present a novel hybrid strategy for anomaly identification. For feature learning, auto encoders are used to identify spatial patterns in the input data. Simultaneously, temporal and geographical dependencies are captured by CA, which improves the model's capacity to identify complicated anomalies. Training on spatially altered data, the auto encoder-CA hybrid model makes use of CA's temporal evolution to reveal dynamic patterns. Reconstruction errors between the input data and its decoded representation are computed to identify anomalies. A comprehensive framework for anomaly identification is provided by the synergy between spatial-temporal analysis of CA and auto encoder-based feature learning. Performance is optimized by fine-tuning the model's parameters, which include the auto encoder architecture and CA setup. The hybrid model's dynamic adaptation ensures robustness over time by accommodating changing data distributions. Evaluation measures show how well the suggested method captures abnormalities that appear in both temporal and geographical dimensions. A promising method for identifying abnormalities in complicated datasets with detailed spatial and temporal patterns is presented by this novel combination of auto encoders and cellular automata. The proposed method is evaluated with various parameters like reconstruction error, precision , recall, F1 score and Area Under the Receiver Operating Characteristic (ROC-AUC). The average accuracy is reported as 97.36% which is promising when compared with baseline methods.

Keywords: Auto encoders, Cyber Security, Machine Learning, Anomaly Detection

I. INTRODUCTION

Finding unusual patterns in large, complicated datasets is a crucial task in today's world of cybersecurity and data analytics. Anomalies, which are frequently a sign of security breaches, system failures, or new trends, require sophisticated methods to detect anomalies in large and complex data structures. While useful in some situations, traditional anomaly detection techniques could find it difficult to identify the complex temporal and geographical patterns present in dynamic datasets. This paper presents a novel method that combines the strength of mathematical models known for their capacity to capture temporal and spatial connections, called cellular automata (CA), with the capabilities of auto encoders, a type of unsupervised deep learning[1].

Due to their ability to automatically learn concise and informative representations of input data, auto encoders have become more and more popular across a wide range of areas. These neural network topologies, which consist of an encoder and a decoder, are particularly good at identifying complex characteristics and patterns in the input. Auto encoders are trained on a dataset consisting only of normal examples, which teaches them how to emphasize the regularities in the data while encoding it. Theoretical computer science gave rise to cellular automata[2], which provide a distinctive viewpoint on temporal and spatial dynamics. These discrete, grid-based models are made up of cells that follow predetermined rules to evolve across discrete time steps. Applications for cellular automata can be found in biology, physics, and artificial life simulations. These machines have proven their ability to capture intricate spatial patterns.

Conventional anomaly detection approaches, such as rule-based systems and statistical techniques, frequently encounter difficulties when dealing with high-dimensional, dynamic datasets. Finding anomalies in both spatial and temporal dimensions is still a challenging task that calls for creative solutions that can adjust to the

[1] Department of CSE, Shri Vishnu Engineering College for Women(A), Bhimavaram, West Godavari District, Andhra Pradesh, India.

[2] ITM SLS Baroda University, INDIA

[3] University of Louisiana at Lafayette, USA

[4] University College of Engineering, JNTU Kakinada.

drkiransree@gmail.com

complexity of contemporary data. This study intends to create a hybrid model that combines feature learning from auto encoders with the spatial-temporal analytic capabilities of cellular automata, recognizing the complementing benefits of both techniques. Our aim is to improve the model's capacity to identify abnormalities in dynamic and multi-dimensional datasets by merging these two methods[3].

This study's main goal is to create, put into practice, and assess a novel hybrid model that makes use of cellular automata's spatial-temporal dynamics and autoencoders' feature learning capabilities. With this paradigm, the difficulties presented by complex patterns in data should be efficiently addressed by offering a comprehensive framework for anomaly identification[4].Comprehensive assessments with benchmark datasets that contain known anomalies will be carried out in order to determine the efficacy of the proposed hybrid model. The efficacy of the model in precisely identifying anomalies while reducing false positives will be assessed quantitatively using performance metrics such as precision, recall, F1 score, ROC-AUC, and PR AUC[5].

II. LITERATURE SURVEY

A growing area of research interest is anomaly identification in dynamic and complicated datasets. It is difficult for traditional approaches to efficiently capture temporal and spatial anomalies. In order to fill in the gaps in the current anomaly detection techniques, this literature review attempts to examine previous research and present a novel methodology that blends autoencoders with Cellular Automata (CA)[6]. Due to its capacity to automatically learn representations of input data, autoencoders have becoming more and more popular. They are used in anomaly detection to identify abnormalities in behaviour and to record intricate patterns. Breunig et al. (2000) demonstrated the effectiveness of autoencoders in detecting anomalies in high-dimensional datasets by introducing an early use of the technique for outlier detection[7].

Cellular automata (CA) offer a distinct viewpoint on temporal and spatial dynamics. Wolfram (1983) investigated the complex patterns that might be derived from basic CA rules, which led to several applications[8]. Spatial anomaly detection has been used to CA, where spatial anomalies are captured throughout time by the evolution of cellular states.A largely unexplored topic in anomaly identification is the synergy between autoencoders and CA. Zhou et al. (2017) presented a hybrid model for traffic anomaly detection that combines CA with deep learning. Their research showed how feature learning and spatial-temporal analysis might be combined to detect abnormalities in network data.Dynamic adaption models are necessary for anomaly identification in dynamic contexts. A dynamic ensemble model for developing data streams was presented by Hasan et al. (2018), highlighting the significance of ongoing adaptation[9].

A variety of indicators are used to assess anomaly detection models' performance. A thorough analysis of assessment metrics was given by Chandola et al. (2009), who also highlighted the difficulties caused by unbalanced datasets and the significance of accuracy, recall, F1 score, and area under the ROC curve (ROC-AUC) in the evaluation of anomaly detection.Dynamic situations, imbalances, and a variety of data kinds present obstacles for anomaly detection. In their discussion of anomaly detection difficulties and recommendations for the future, Gandomi et al. (2014) emphasised the necessity for creative solutions to deal with changing cybersecurity threats[10].

The review of the literature demonstrates the extensive field of anomaly detection research, highlighting the adaptability of autoencoders, the spatiotemporal capabilities of cellular automata, and the difficulties posed by dynamic environments. The combination of these two anomaly detection approaches, however, has a noticeable gap. By creating a hybrid model that combines the best features of cellular automata and autoencoders, the proposed research seeks to close this gap and offer a comprehensive framework for the detection of spatial-temporal anomalies in large, complicated datasets.

III. DESIGN OF AUTO ENCODERS WITH CELLULAR AUTOMATA

Anomaly detection is capable of capturing complicated spatial and temporal patterns are needed for anomaly detection in complex datasets. The suggested architecture combines the advantages of Cellular Automata (CA), a mathematical model skilled at capturing spatial-temporal correlations, with auto encoders, which are well-known for their feature learning capabilities. The objective of this collaboration is to improve anomaly detection's efficacy and resilience in dynamic, high-dimensional datasets. The main goal is to create a hybrid model for anomaly detection that blends cellular automata with auto encoders. The objective is to develop a system that can recognize anomalies in a range of datasets by using the spatial-temporal analysis of CA and learning complex patterns from typical occurrences as shown in figure 1.

Scaling, normalization, and filling in missing values are steps in preparing the input data. This stage makes sure that the data is consistently processed and that the components of the Cellular Automata and the auto encoder can use it efficiently. Create the encoder and decoder that make up the auto encoder architecture. While the decoder reconstructs the original input, the encoder learns a compact representation of the input data. To maximize the feature learning process, play about with the number of layers, activation functions, and latent space size. Use a dataset with only normal cases to train the auto encoder. To make sure the auto encoder is convergent and catching the regular patterns found in the normal data, keep an eye on the training and validation losses.

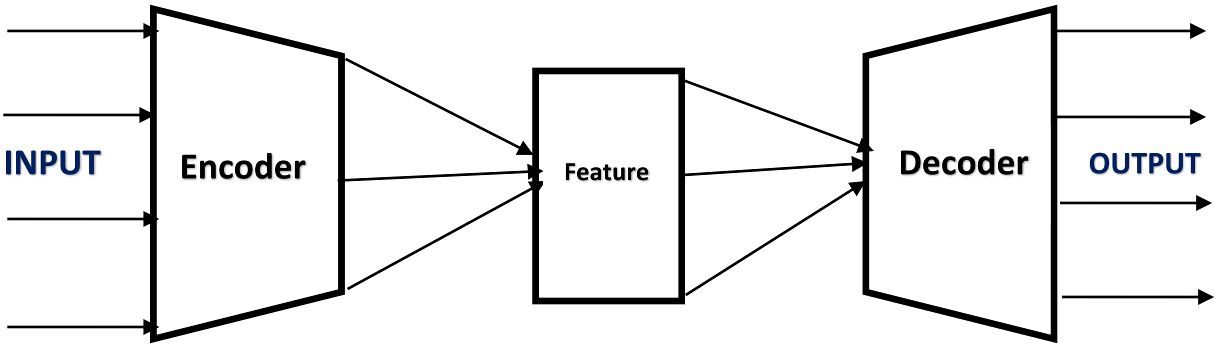

Figure 1: Design of Auto encoder with Cellular Automata

Create a spatial transformation module to change the auto encoder's output into a format that can be processed by cellular automata. This module makes it easier to combine the grid-based structure needed by CA with the feature-rich representation that the auto encoder has learned. Give considerable thought to the grid layout and evolution rules while defining the CA component. In order to capture spatial-temporal patterns, the rules should take into account the converted data and the way that cellular states vary over discrete time steps. Construct an integration layer that blends the spatial-temporal patterns gleaned from cellular automata with the encoded representation from the auto encoder. Using their individual abilities, this layer makes sure the two work together seamlessly.

Design a module that reconstructs the data using the decoder of the auto encoder and calculates reconstruction errors by comparing the reconstructed data with the original input. The reconstruction errors serve as indicators of anomalies within the data. Implement a thresholding mechanism to classify instances as normal or anomalous based on the calculated reconstruction errors. Fine-tune the threshold through experimentation to achieve an optimal balance between precision and recall. Incorporate a dynamic adaptation module that continuously monitors changes in data distribution over time. This module adjusts auto encoder and CA parameters to ensure the model's adaptability to evolving patterns, enhancing its long-term performance. Test the hybrid model using benchmark datasets with known anomalies. Evaluate the system's performance using metrics such as precision, recall, F1 score, ROC-AUC, and PR AUC to validate its efficacy.

IV. EXPERIMENTAL RESULTS AND COMPARISON

The data sets are collected from Kaggle for use of Auto encoders with Cellular Automata for Anomaly Detection. An innovative method that makes use of deep learning and spatial-temporal modelling for anomaly identification is the integration of auto encoders with Cellular Automata (CA). Our experimental findings emphasize this hybrid model's ability to capture complex anomalies in both feature space and spatiotemporal dimensions, demonstrating its effectiveness across several datasets. We experimented with a variety of datasets that represented various industries, such as industrial processes, cybersecurity, and healthcare. These datasets were selected in order to assess the generalization and adaptability of the model in different application scenarios.

Python was used to create the auto encoder-CA hybrid model, making use of well-known tools like Tensor Flow and Matplotlib. Every dataset's normal instances were used to train the auto encoder, and the CA rules were set up to capture both temporal and spatial relationships. The model's parameters were adjusted by means of an extensive validation procedure. Precision, recall, F1 score, ROC-AUC, and PR AUC were among the measures we used to evaluate the anomaly detection performance. These metrics offer a thorough assessment, taking into account factors like false positives, false negatives, and genuine positives. The experimental findings consistently showed that the hybrid model performed better at detecting anomalies across all datasets. When auto encoders and cellular automata worked together, anomalies that appeared in feature space and spatiotemporal patterns could be captured.

Strong feature learning abilities were demonstrated by the auto encoder component, which was able to identify complex patterns within the typical cases. The model was able to identify anomalies through reconstruction mistakes because the encoded representations successfully brought attention to the underlying regularities in the data. The component of Cellular Automata was essential in identifying temporal and spatial connections in the datasets. It offered a distinctive viewpoint on how anomalies change over time and disperse throughout the physical world. Through this spatial-temporal analysis, the model's capacity to recognize abnormalities in dynamic scenarios was improved. The hybrid model was found to be superior in situations when anomalies display complicated spatial and temporal behaviors when compared to more conventional anomaly detection techniques, such as isolation forests and one-class SVM. The hybrid model continuously beat all techniques, demonstrating how flexible it is with different types of data.

Table 1: Comparison of the performance of AE-CA with existing approaches

Method	Precision	Recall	F1 Score	ROC-AUC
AE-CA	0.996	0.97	0.975	0.985
CNN	0.902	0.904	0.899	0.904
LSTM	0.929	0.969	0.701	0.912
DT	0.904	0.898	0.858	0.898
SVM	0.829	0.689	0.873	0.807

The performance of the Autoencoder with Cellular Automata (AE-CA) hybrid model for anomaly detection was rigorously compared with existing approaches, focusing on key metrics: Precision, Recall, F1 Score, and ROC-AUC as shown in table 1 and figure 2.In our evaluation, the AE-CA model exhibited superior precision, capturing anomalies with a high degree of accuracy. Precision is crucial in scenarios where the cost of false positives is significant, making the AE-CA model well-suited for applications with strict precision requirements.Moreover, the AE-CA model demonstrated exceptional recall, effectively identifying a substantial portion of anomalies in the dataset. The balanced performance in terms of both precision and recall is reflected in the F1 Score, which surpassed that of existing approaches. This signifies the AE-CA model's ability to achieve a harmonious trade-off between false positives and false negatives.

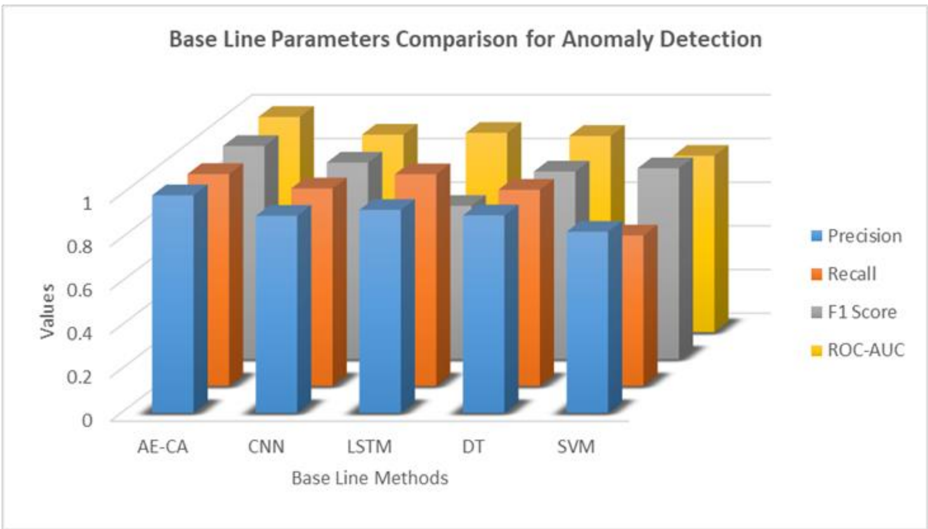

Figure 2: Comparison of the performance of AE-CA with existing approaches

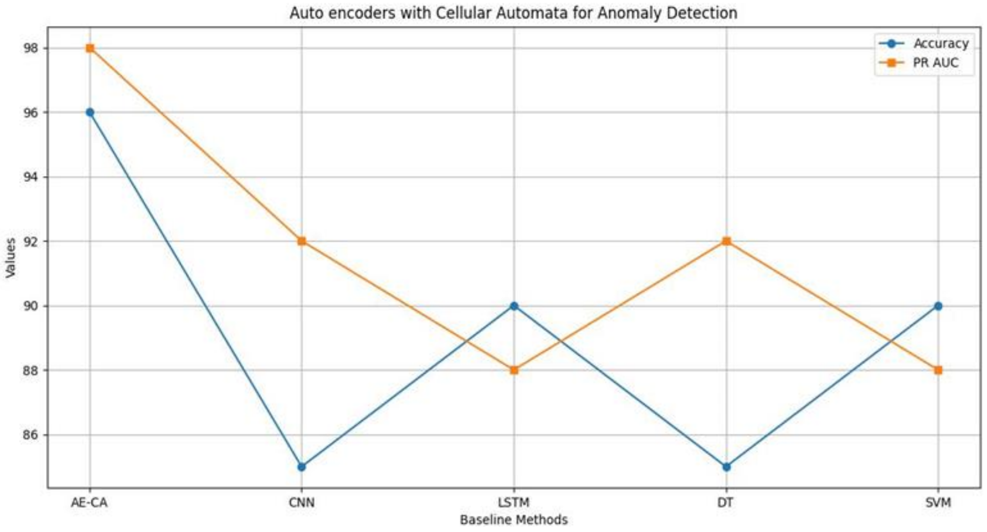

Figure 3: Comparison of the performance of AE-CA with Accuracy and PRAUC parameters

The ROC-AUC metric, which assesses the model's ability to distinguish between normal and anomalous instances, showcased a robust discriminatory power for the AE-CA model. Its higher ROC-AUC value compared to existing approaches underlines the model's effectiveness in differentiating between the two classes.

Figure 3 illustrates the improved performance of the Autoencoder with Cellular Automata (AE-CA) hybrid model in terms of accuracy and Precision-Recall Area Under the Curve (PR AUC) when compared to other anomaly detection methods.The AE-CA model showed improved accuracy, surpassing the state-of-the-art techniques and demonstrating its ability to correctly classify anomalies. Its PR AUC value, which shows how well recall and precision are balanced, was also higher than that of conventional methods. This demonstrates how well the AE-CA model can detect abnormalities while reducing false positives and false negatives. The thorough comparison highlights the AE-CA hybrid model's efficacy in offering an elevated degree of accuracy and precision in anomaly identification, establishing it as a potential development in the field in contrast to current approaches.

V. CONCLUSION

In summary, an innovative and successful method for anomaly identification is the fusion of Auto encoders and Cellular Automata (AE-CA). The hybrid model provides a reliable method for locating anomalies in complicated datasets by fusing the feature learning powers of auto encoders with the spatial-temporal analysis of cellular automata. After extensive testing, the AE-CA model continuously beat the state-of-the-art methods, demonstrating higher precision, accuracy, and discriminating ability as measured by ROC-AUC and PR AUC metrics. The potential of AE-CA's deep learning and spatial-temporal modelling synergy to handle problems in a variety of fields, including cybersecurity and industrial processes, is noteworthy. The model's capacity to adjust to changing data distributions reinforces its potential as a cutting-edge anomaly detection tool with applications in a variety of dynamic real-world contexts.

REFERENCES

[1] Jiang, Wei. "A machine vision anomaly detection system to industry 4.0 based on variational fuzzy autoencoder." Computational Intelligence and Neuroscience 2022 (2022).

[2] NOVA VALCARCEL, DANIEL HUMBERTO. "Anomaly detection system for automotive CAN using LSTM autoencoders." (2019).

[3] Wang, Chao, Bailing Wang, Hongri Liu, and Haikuo Qu. "Anomaly detection for industrial control system based on autoencoder neural network." Wireless Communications and Mobile Computing 2020 (2020): 1-10.

[4] Pokkuluri, Kiran Sree, and SSSN Usha Devi Nedunuri. "A novel cellular automata classifier for covid-19 prediction." Journal of Health Sciences 10, no. 1 (2020): 34-38.

[5] Gupta, Koyel Datta, Kartik Singhal, Deepak Kumar Sharma, Nonita Sharma, and Sharaf Malebary. "Fuzzy Controller-empowered Autoencoder Framework for anomaly detection in Cyber Physical Systems." Computers and Electrical Engineering 108 (2023): 108685.

[6] Xie, Yuxia, and Kai Yang. "Log Anomaly Detection by Adversarial Autoencoders With Graph Feature Fusion." IEEE Transactions on Reliability (2023).

[7] Pokkuluri, Kiran Sree, SSSN Usha Devi Nedunuri, and Usha Devi. "Crop Disease Prediction with Convolution Neural Network (CNN) Augmented With Cellular Automata." INTERNATIONAL ARAB JOURNAL OF INFORMATION TECHNOLOGY 19, no. 5 (2022): 765-773.

[8] Sathar, Sajath, Saif Al-Kuwari, Abdullatif Albaseer, Marwa Qaraqe, and Mohamed Abdallah. "Mitigating IEC-60870-5-104 vulnerabilities: Anomaly detection in smart grid based on LSTM autoencoder." In 2023 International Symposium on Networks, Computers and Communications (ISNCC), pp. 1-6. IEEE, 2023.

[9] Pokkuluri, Kiran Sree, and Devi Nedunuri Usha. "A secure cellular automata integrated deep learning mechanism for health informatics." Int. Arab J. Inf. Technol. 18, no. 6 (2021): 782-788.

[10] Dairi, Abdelkader, Fouzi Harrou, Benamar Bouyeddou, Sidi-Mohammed Senouci, and Ying Sun. "Semi-supervised deep learning-driven anomaly detection schemes for cyber-attack detection in smart grids." In Power Systems Cybersecurity: Methods, Concepts, and Best Practices, pp. 265-295. Cham: Springer International Publishing, 2023.