



Article

Predicting Cryptocurrency Fraud Using ChaosNet: The Ethereum Manifestation

Anurag Dutta ¹, Liton Chandra Voumik ^{2,*}, Athilingam Ramamoorthy ³, Samrat Ray ⁴ and Asif Raihan ⁵

¹ Department of Computer Science, Government College of Engineering and Textile Technology, Serampore, Calcutta 712201, India; anuragdutta.research@gmail.com

² Department of Economics, Noakhali Science and Technology University, Noakhali 3814, Bangladesh

³ Department of Mathematics, Velammal Engineering College, Anna University, Chennai 600066, India

⁴ The Institute of Industrial Management, Economics and Trade, Peter the Great St. Petersburg Polytechnic University, St. Petersburg 190005, Russia

⁵ Institute of Climate Change, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia

* Correspondence: litonvoumik@gmail.com

Abstract: Cryptocurrencies are in high demand now due to their volatile and untraceable nature. Bitcoin, Ethereum, and Dogecoin are just a few examples. This research seeks to identify deception and probable fraud in Ethereum transactional processes. We have developed this capability via ChaosNet, an Artificial Neural Network constructed using Generalized Luröth Series maps. Chaos has been objectively discovered in the brain at many spatiotemporal scales. Several synthetic neuronal simulations, including the Hindmarsh–Rose model, possess chaos, and individual brain neurons are known to display chaotic bursting phenomena. Although chaos is included in several Artificial Neural Networks (ANNs), for instance, in Recursively Generating Neural Networks, no ANNs exist for classical tasks entirely made up of chaoticity. ChaosNet uses the chaotic GLS neurons' property of topological transitivity to perform classification problems on pools of data with cutting-edge performance, lowering the necessary training sample count. This synthetic neural network can perform categorization tasks by gathering a definite amount of training data. ChaosNet utilizes some of the best traits of networks composed of biological neurons, which derive from the strong chaotic activity of individual neurons, to solve complex classification tasks on par with or better than standard Artificial Neural Networks. It has been shown to require much fewer training samples. This ability of ChaosNet has been well exploited for the objective of our research. Further, in this article, ChaosNet has been integrated with several well-known ML algorithms to cater to the purposes of this study. The results obtained are better than the generic results.

Keywords: Cryptocurrency; blockchain; ChaosNet; GLS Neurons; Artificial Neural Network



Citation: Dutta, Anurag, Liton Chandra Voumik, Athilingam Ramamoorthy, Samrat Ray, and Asif Raihan. 2023. Predicting Cryptocurrency Fraud Using ChaosNet: The Ethereum Manifestation. *Journal of Risk and Financial Management* 16: 216. <https://doi.org/10.3390/jrfm16040216>

Academic Editor: Thanasis Stengos

Received: 11 February 2023

Revised: 11 March 2023

Accepted: 27 March 2023

Published: 29 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Learning through techniques such as Machine Learning (ML) and Deep Learning is now possible, thanks to the development of Artificial Intelligence (Lauriola et al. 2022). These methods have gained appeal, with applications in practically every sector of human endeavors. Among these are, to name a few, voice processing (Gray 2009), computer vision (Jiao et al. 2019), cyber security (Kianpour et al. 2021), and medical diagnosis (Bhavsar et al. 2021). Despite being influenced by the biological brain, humans' learning and memory encoding processes are not directly tied to these algorithms. The teaching processes for changing masses and biases in these Artificial Neural Networks (ANNs) (Jospin et al. 2022) are standardized based on optimization strategies and the principle of minimizing loss and error functions. As a larger pool of nascent data is fed into the system, the ANNs currently use a considerable amount of subjected hyperparameters (Yang and Shami 2020) fixed via ad hoc approaches to achieve better prediction. These synaptic alterations are based primarily on outcomes and lack or have little solid theoretical

support. Additionally, these methods require a considerable quantity of training data to forecast or classify the target classes' distribution accurately.

ANNs have succeeded in their defined tasks but must catch up to the human intellect when completing tasks such as natural language processing (Turchin and Builes 2021). Researchers are concentrating on creating biologically inspired algorithms and architectures to utilize the remarkable learning capabilities of the Homo sapiens brain while contributing to a better understanding of the brain. Primary areas of focus include memory encoding and learning. One of the brain's most intriguing traits is its capacity for "chaos"—the phenomenon whereby straightforward deterministic nonlinear systems exhibit complex, unexpected, and seemingly random behavior. Electroencephalogram (EEG) signals (Montoya-Martínez et al. 2019) are known to have chaotic dynamics (Shen et al. 2021). A neural system's sensitivity to small changes in internal functioning characteristics aids in producing the optimal response to various influences. This characteristic resembles the chaotic system's dynamic features. Furthermore, the brain constantly switches between several states, rather than returning to homeostasis after a transient change. For this reason, it is hypothesized that the brain can display various behaviors, including periodicity in orbits, a weak nature of chaotic dynamics, or a strong nature of chaos, depending on the functional parameters of the neurons. Cerebral networks, which are made up of trillions of neurons, exhibit chaotic activity, but individual neurons at the cellular and subcellular levels also display these dynamics. These neurons' ability to build impulse trains allows the brain to transmit and store information. When various ions pass across the axonal membrane and affect the voltage, action potentials or impulses are produced. Regarding the communication bridging the ion passages and the axonal membrane, Huxley and Hodgkin initially put forth a dynamic system model that can create real action potentials (Hodgkin and Huxley 1952). Later, it was suggested that neural networks use its streamlined counterparts, such as the Hindmarsh–Rose (Hindmarsh and Rose 1984) and the FitzHugh–Nagumo models (FitzHugh 1961; Nagumo et al. 1962). These models all display chaotic behavior.

Recurrent neural networks (Moses et al. 2021; Hewamalage et al. 2021) are one type of artificial neural network that exhibits chaotic dynamics; however, as far as we know, none of these proposed architectures thus far demonstrate chaos at the level of individual neurons when subjected to classification tasks. Other chaotic neuron models have been proposed as a theoretical description of brain memory encoding.

One of these models is the Aihara model (Aihara et al. 1990), which has been applied to cognitive tasks in the network's erratic periodic orbits (Crook and Scheper 2008). Freeman, Kuzma, and their colleagues developed chaotic simulations motivated by the mammalian sensory pathways to demonstrate the process of memorizing scents (Meurant 2012; Chang and Freeman 1996; Kozma and Freeman 1999). Chaos in neural networks has also been studied by Tsuda and others. Globally coupled chaotic maps' dynamic properties have been reviewed by Kaneko, who hypothesized that these networks would be able to handle biological data.

Generalized Luröth Series (GLS) 1D chaotic map neurons make up ChaosNet (see Figure 1), an artificial neural network (ANN) (Harikrishnan and Nagaraj 2019). This network can learn from a few training examples to perform classification tasks. ChaosNet was developed to utilize some of the best characteristics of biological neural networks. It has been demonstrated that, while using significantly fewer training samples than traditional ANNs, it can perform complex classification tasks on par with or better than conventional ANNs.

ChaosNet, inspired by biological neurons, uses a property similar to the "spike-count rate" of the firing of chaotic neurons as a neural code for learning (see Figure 2). Additionally, the network can exhibit a hierarchical architecture, incorporating information as it is transmitted to deeper, higher levels of the network. A Generalized Luröth Series, or GLS, is a piecewise linear 1D chaotic map representing the neuron we specify. Examples of

GLS include the well-known tent map, the binary map, and the tent map's skewed relatives. The sorts of GLS neurons that are employed in ChaosNet are:

$$T_{Skew-Binary}(x) = \begin{cases} \frac{x}{b} & 0 \leq x < b \\ \frac{(x-b)}{(1-b)} & b \leq x < 1 \end{cases} \quad (1)$$

and

$$T_{Skew-Tent}(x) = \begin{cases} \frac{x}{b} & 0 \leq x < b \\ \frac{(1-x)}{(1-b)} & b \leq x < 1 \end{cases} \quad (2)$$

A cryptocurrency ([Makarov and Schoar 2019](#)), often called a crypto-currency or just a “crypto,” is digital money supported or maintained by no single central body, such as a bank or government. It is a decentralized means of verifying that the parties to a transaction genuinely have the funds they claim. It eliminates the need for traditional intermediaries such as banks when money is transferred between two businesses. Digital ledgers are computerized databases that use safe encryption to protect transaction records, regulate the production of new currencies, and confirm ownership transfers, which are used to maintain individual coin ownership records. Cryptocurrency is typically not authorized by a centralized unit and does not exist in a tangible form like paper money. In contrast to digital currencies managed by a central bank, cryptocurrency usually employs decentralized control (CBDC). When a cryptocurrency ([Goodell and Aste 2019](#)) is coined, generated in anticipation of issuance, or released by a single issuer, it is considered to be centralized. When utilized with decentralized governance, each cryptocurrency uses distributed ledger technology, generally a blockchain, which acts as a public database of financial transactions. Currency, commodities, and stocks are traditional asset classes and macroeconomic indicators with moderate sensitivity to cryptocurrency returns.

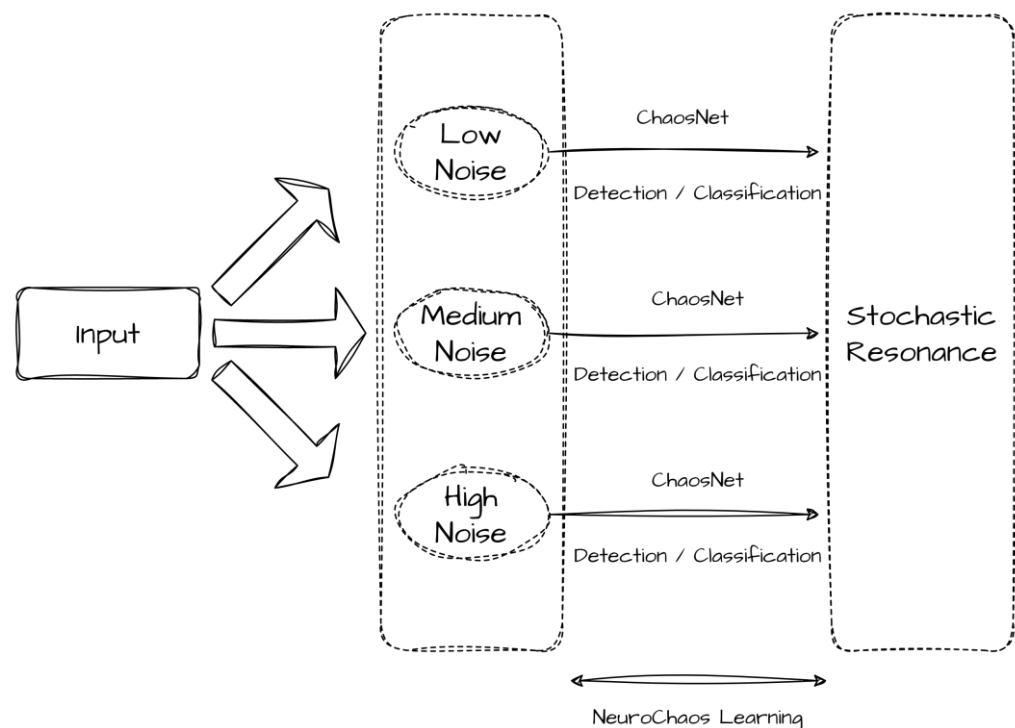


Figure 1. Neurochaos Learning ([Harikrishnan and Nagaraj 2021](#)).

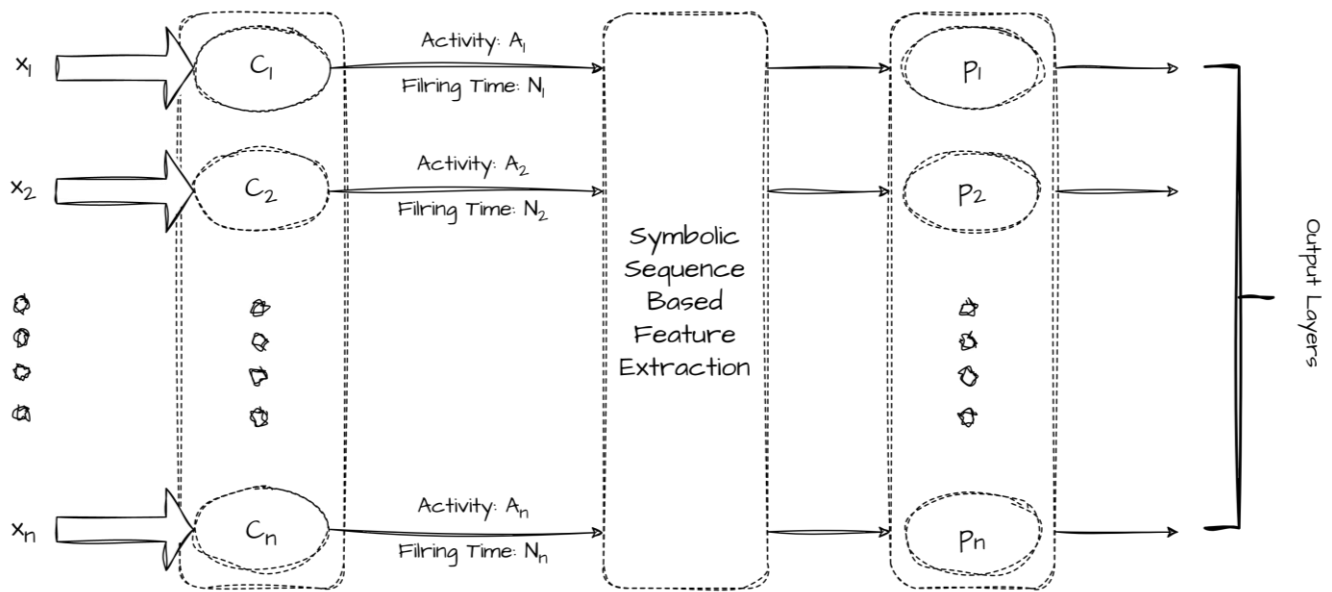


Figure 2. The architecture of ChaosNet (Balakrishnan et al. 2019) Luröth neural networks for purposes relating to classification. C_1, C_2, \dots, C_n are the unit dimensional GLS neurons. Each neuron initially exhibits q units of normalized neuronal activity. The input to the network, or the normalized collection of stimuli, is denoted by the $\{x_i\}_{i=1}^n$. When a GLS neuron C_i 's chaotic activity value $A(t)$, starting from initial neural activity (q), reaches the ε -a neighborhood of stimulus, it stops firing chaotically. This neuron has a “firing time” of N_i ms. $A(t)$ contains topological transitivity symbolic sequence feature p_i . This feature is extracted from the $A(t)$ of the C_i 's GLS-neuron.

Financial or personal gain is the intended outcome of cryptocurrency fraud, which is dishonest behavior in the cryptocurrency business; by convincing their unwitting victims to take an action, such as clicking on a link or disclosing personal information, scammers and hackers on the internet hope to make some fast money.

For cryptocurrency scams, criminals frequently try to gain access to a victim's digital wallet to steal their cryptocurrency assets. Typically, they will ask a victim to connect their wallet to a bogus website or deceive the victim into giving them access to their wallet's private keys. Cryptocurrency fraud can be of many types, but, broadly, it can be categorized into the following four types:

1. **Phishing:** Although fraudsters are nothing new, individuals continue to fall victim to this tactic. A malicious hyperlink in an inbox or a fraudulent website that occasionally uncannily resembles its genuine counterpart can be used in phishing scams. A victim's personal information, such as their internet passwords or the private keys to their crypto wallet, may be stolen using the link or website.
2. **Man-in-the-Middle:** Man-in-the-middle assaults are a technique that con artists use to obtain personal information, much like phishing scams. To access a victim's bitcoin wallet or private account information, a fraudster will disrupt a Wi-Fi session on a broad network instead of doing so through links. One can use a VPN to secure their data while depositing cryptocurrency to avoid this.
3. **Investment Scam:** Investment managers who offer to help a person make significant improvements to their portfolio may be fraudsters. They will entice customers to transmit their cryptocurrencies and may even promise to increase the value of their investments by 50 times. Forbes Advisor does caution that “if you comply with their demands, kiss goodbye to your cryptocurrency.” Using this scam, the con artist probably deceives several people, takes their cryptocurrency, and then vanishes.
4. **Pump-and-Dump:** This is a tactic used in both regular stock markets and cryptocurrency marketplaces. When a coin launches, its owners sell all their holdings, known as a pump-and-dump strategy. As a result, the price reaches an erroneous peak before

dropping sharply after the initial public offering is over. False statements made about a project that cause a lot of hype can worsen the impact of these tactics.

2. Ethereum

Ethereum (Tikhomirov 2018) is moving to a proof-of-stake consensus algorithm but it was not deployed initially as such. It is well known for its Ether cryptocurrency (ETH). Anyone can use Ethereum to develop safe digitizing systems. It has a currency designed to reimburse users for work done in support of the blockchain, but if accepted, users may also use it to exchange for physical goods and services. Ethereum has the characteristics of being extensible, adaptable, anonymous, and decentralized. It is the decentralized cryptocurrency of choice for programmers and businesses, which has led to building technology based on Ether and altered multiple industries and how people go about their daily lives. In late 2013, Vitalik Buterin, a developer and cofounder at Bitcoin Magazine, introduced Ethereum (Atzei et al. 2017) as a mechanism for building decentralized apps in a white paper. Buterin told the Bitcoin Kernel technicians that applications other than currency may be derived from the nature of blockchain technology and suggested that a more sophisticated language for designing apps was needed. In early 2014, Ethereum Switzerland GmbH, a Swiss corporation, began officially developing the software underpinning Eth-Suisse (Bhargavan et al. 2016). The concept of holding executable intelligent contracts on the blockchain had to be outlined before it was implemented in software. This work was done in the Ethereum Virtual Machine specification by Gavin Wood, the Ethereum Yellow Paper's then-Chief Technical Officer. The Stiftung Ethereum (Bentov et al. 2016) (Ethereum Foundation) was established as a Swiss non-profit organization. From July through August 2014, an online public crowd sale was held in which people bought the Ethereum value token (ether) with bitcoin, another digital money. Although Ethereum's technical advances were first lauded, concerns were raised about its scalability and security. To construct and achieve consensus on an ever-expanding collection of "blocks," or groups of transactions known as a blockchain, Ethereum is an epicondyle (Trusted Smart Contracts 2017), or virtual collective (Sompolinsky and Zohar 2015), of computer nodes. Each block has a distinct identifier for the sequence that must come before each block to be considered authentic. When a base station adds a block to its chain, it executes the actions in the block in the designated order, each of which can potentially alter the ETH balance (Chen et al. 2017) and other rack values of Ethereum accounts. In a Merkle tree, the "state," or collection of these totals and values, is held on the node apart from the blockchain. Only a limited portion of the network, known as its "peers," are accessible to each node. Every time a node wants to add a new transaction (Filliâtre and Paskevich 2013; Dutta et al. 2022) to the chain, it sends copies of the transaction to all of its contemporaries, who then send copies to all of their contemporaries, and so forth. It spreads throughout the network in this way. All of these new transactions are tracked by a group of nodes known as miners, who use them to build new blocks and distribute them to the remainder (Choudhury and Dutta 2022) of the network. Every time a node receives a partnership, it verifies the validity of the block and each transaction contained inside. If the block is valid, it is added to the blockchain, and each transaction is carried out. A node may receive numerous blocks vying to succeed a specific block, since block generation (Metcalf 2020) and broadcasting are permissionless. The node records each valid chain that results from this and routinely discards the shortest one: the Ethereum protocol (Decker and Wattenhofer 2013) states that the longest chain is to be considered at any given time.

3. Data and Methods

3.1. Dataset Description

We have collected a set of Ethereum transaction details using the Etherscan API and etherscan API. The dataset has 14 features, namely,

- Avgminbetweensenttnx: Minutes between each transaction on average for the account.

- Avgminbetweenreceivedtnx: Minutes between transactions received on average for the account.
- TimeDiffbetweenfirstand_last(Mins): Minutes between the first and last transactions.
- Sent_tnx: Total volume of typical transactions sent.
- Received_tnx: Total volume of typical transactions received.
- NumberofCreated_Contracts: Total number of contract transactions created.
- UniqueReceivedFrom_Addresses: Total unique addresses from which transactions were sent to the account.
- UniqueSentTo_Addresses: Total unique addresses to which transactions were sent from the account.
- MinValSent: Lowest amount of Ether sent.
- MaxValSent: Highest amount of Ether sent.
- AvgValSent: Average amount of Ether sent over time

The dataset is available at https://github.com/Anurag-Dutta/Ethereum/blob/19b35453da25b40bb22556c1070cfb79fbb52b2f/Eth_Pub_19122022.csv (accessed on 26 March 2023), which was churned from the open-source database, <https://github.com/MrLuit/EtherScamDB> (accessed on 26 March 2023) and integrated by <https://etherscan.io> (accessed on 26 March 2023).

Since cryptocurrency transactions cannot be easily traced, the dataset that we have snipped from the Etherscan API and etherscamdb API will not be sufficient for any ML classifiers to classify them as being fraudulent or not. However, ChaosNet is well-known for its ability to be trained from very few data instances or very few data points. Further, the snipped data had numerous features inside of them. Some of these features, such as Index, Address, etc., could be more useful for prediction than others. Thus, we subjected them to the PCA functionality of the decomposition utility of the sklearn package. More details regarding the decomposition utility can be referred to using the sklearn documentation. Finally, the models were made to run on 10 columns. However, the reduction in dimension is not meant to be used as an object. This reduction value can be any number between six and 14. In our work, the features were selected based on the information gained from each of them. Gaining knowledge from a random variable X as determined by a random variable observation χ is defined as χ taking value $\chi = \zeta$ as

$$\Delta I_{X,\chi}(X, \zeta) = \mathfrak{D}_{kullback-leibler} \left(\frac{P_X\left(\frac{x}{\zeta}\right)}{P_X\left(\frac{x}{I}\right)} \right) \quad (3)$$

where

$P_X\left(\frac{x}{I}\right)$ is the Prior Distribution concerning Kullback–Leibler divergence and $P_X\left(\frac{x}{\zeta}\right)$ is the Posterior Distribution concerning Kullback–Leibler divergence. The higher the value of $\Delta I_{X,\chi}(X, \zeta)$, the better is the knowledge gain. Therefore, we hope to obtain better results by using this formula.

3.2. Methods: ChaosFeatureExtractor + ML Classifiers

Using ChaosNet Standalone, good performance was achieved compared to the classic machine learning classifiers. Nonetheless, we can achieve better results by using a better ML classifier in conjunction with the Chaos Feature Extractor (see Figure 3) (Sethi et al. 2023).

ChaosNet uses three hyperparameters:

INA—Initial Neural Activity
EPSILON_1—Noise Intensity
DT—Discrimination Threshold

This Single Internal Neuron's memory corresponds to the Initial Neural Activity. We have used AdaBoost and kNN (k-nearest neighbors) as individual machine learning classifiers. We have also made use of ChaosNet Standalone.

The respective values of the hyperparameters for the same were tuned to:

INITIAL_NEURAL_ACTIVITY = [0.38]
 DISCRIMINATION_THRESHOLD = [0.06]
 EPSILON = [0.29]

For Standalone ChaosNet. For ChaosNet Feature Extractor conjugated with AdaBoost, they were tuned to:

INITIAL_NEURAL_ACTIVITY = [0.36]
 DISCRIMINATION_THRESHOLD = [0.06]
 EPSILON = [0.29]

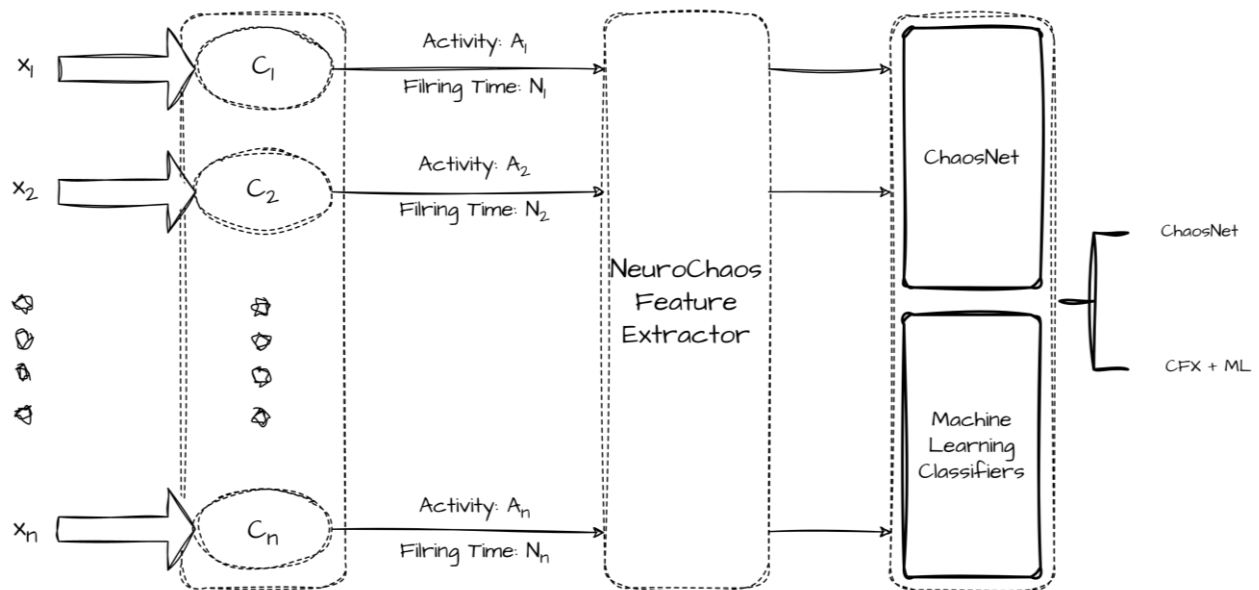


Figure 3. Architecture proposing Conjunction of the Chaos Feature Extractor with standard ML classifiers. The three actions involved include feature transformation, feature extraction from neurochaos, and classification in the first two steps. ChaosNet or any other ML classifier could be selected as the chosen classifier. One-dimensional Generalized Lüroth Series (GLS) neurons make up the initial tier of the feature transformation unit. These are tiny-bit linear chaotic maps. More details regarding this architecture can be obtained from the works by Deeksha Sethi, Nithin Nagaraj, and Harikrishnan N.B. (Sethi et al. 2023).

In 1995, Yoav Freund and Robert Schapire created AdaBoost, a statistical classification meta-algorithm. For their efforts, they received the 2003 Gödel Prize. Combining this with a variety of other learning approaches can improve its performance. The findings of the different learning algorithms, or “weak learners”, are combined to produce a weighted sum representing the boosted classifier’s outcomes. Although AdaBoost can be applied to a wide range of classes or limited intervals on the natural line, it is most often employed for binary classification. AdaBoost is adaptive in that it modifies succeeding weak learners in favor of examples incorrectly identified by earlier classifiers. In some cases, it may be less prone to overfitting than other learning methods. The final model converges to a strong learner even if each learner’s performance is just marginally better than random guessing.

For ChaosNet Feature Extractor conjugated with k-nearest neighbors, the parameters were:

INITIAL_NEURAL_ACTIVITY = [0.039]
 DISCRIMINATION_THRESHOLD = [0.070]
 EPSILON = [0.023]

One of the simplest supervised learning-based nonparametric machine learning algorithms is k-nearest neighbors. Assuming that new cases and data are similar to existing topics, this classifies new cases most similar to existing ones, stores all the relevant data,

and places new data into categories based on similarity. Therefore, it is simple to categorize new data into appropriate categories using the kNN method. Although kNN algorithms can be applied to classification and regression problems, they are most frequently utilized for classification issues. In other words, no presumptions regarding the underlying data are made. It is also known as a delayed learning algorithm since it saves the dataset and modifies it during classification rather than instantly learning from the training set. The kNN algorithm only draws from the training phase dataset and classifies fresh data in the same way as the new data as it comes in.

To evaluate performance, we used the macro F1 score (see Tables 1 and 2). The F1 score can be conceived as a harmonic mean of precision and recall, where one is the best and zero is the worst. Precision and recall are both equally crucial in determining the F1 score; “macro” computes the measurements for each label and derives their unweighted mean. Label imbalance is not considered in this. The confusion matrix is used to calculate this measure. Mathematically,

$$\text{Macro F1 Score} = \frac{F1 \text{ Score}_{\text{Class } 1} + F1 \text{ Score}_{\text{Class } 2} + \dots + F1 \text{ Score}_{\text{Class } n}}{n} \quad (4)$$

$$\therefore \text{Macro F1 Score} = \frac{1}{n} \left(\sum_{i=1}^n F1 \text{ Score}_{\text{Class } i} \right) \quad (5)$$

where

$$F1 \text{ Score}_{\text{Class } i} = \left(\frac{2 \times \text{Precision}_{\text{Class } i} \times \text{Recall}_{\text{Class } i}}{\text{Precision}_{\text{Class } i} + \text{Recall}_{\text{Class } i}} \right)$$

$$\text{Precision}_{\text{Class } i} = \left(\frac{\text{True Positive}_{\text{Class } i}}{\text{True Positive}_{\text{Class } i} + \text{False Positive}_{\text{Class } i}} \right)$$

$$\text{Recall}_{\text{Class } i} = \left(\frac{\text{True Positive}_{\text{Class } i}}{\text{True Positive}_{\text{Class } i} + \text{False Negative}_{\text{Class } i}} \right)$$

Table 1. Comparison metrics for the algorithms used in the article based on their macro F1 score when they were subjected to training.

Algorithm	Macro F1 Score (Training)
ChaosNet Standalone	0.5802753655203908
Chaos Feature Extractor + AdaBoost	0.8125910159305623
Chaos Feature Extractor + kNN	0.7937217353400664

Table 2. Comparison metrics for the algorithms used in the article based on their macro F1 score when they were subjected to testing.

Algorithm	Macro F1 Score (Testing)
ChaosNet Standalone	0.5752543039000217
Chaos Feature Extractor + AdaBoost	0.6649360740269832
Chaos Feature Extractor + kNN	0.7888128840520701

Figures 4–6 denote the confusion matrix of each of these algorithms while subjected to testing on the dataset mentioned in Section 3.1. In the training dataset, they were classified using knowledge from the transactions based on feedback from the sender. The classification was done based on reviews of the crypto transactions. For testing, the model randomly chose 20% of the data from the given data pool. The top-left sector in the confusion matrix represents the cases when, according to the data description, the transaction was considered to be a genuine transaction and the devised model predicted the same, while the bottom-right sector in the confusion matrix represents the cases when,

according to the data description, the transaction was considered to be a fraudulent one and the devised model predicted it to be fraudulent.

For the confusion matrix, True—True sector and False—False sector are the important areas which delineate the performance of any model. In Figure 6, the confusion matrix shows the maximal magnitude for these two sectors. Thus, it is evident that the model using the Chaos Feature Extractor complemented with k-nearest neighbors performed well.

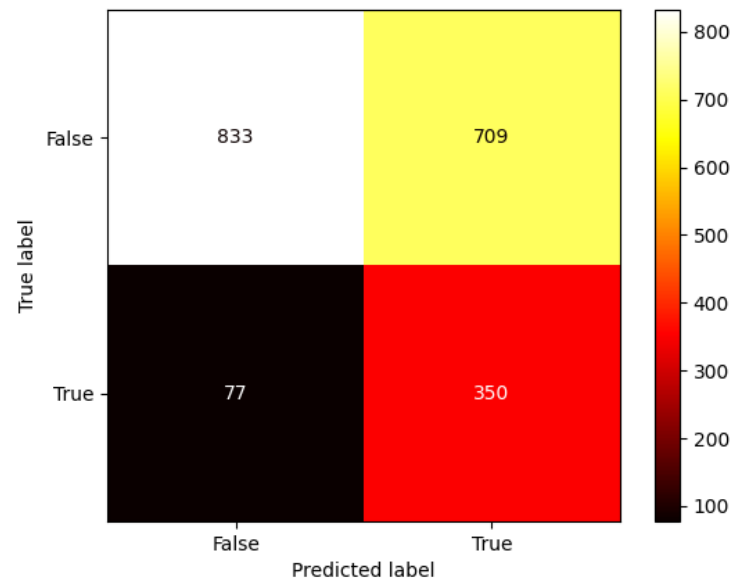


Figure 4. Pictorial representation of the confusion matrix subjected to Ethereum Fraudulence Prediction with detailed annotations. The Standalone ChaosNet suspected 350 transactions to be fraudulent which were in fact fraudulent. The algorithm declared 833 transactions to be genuine which were in fact genuine.

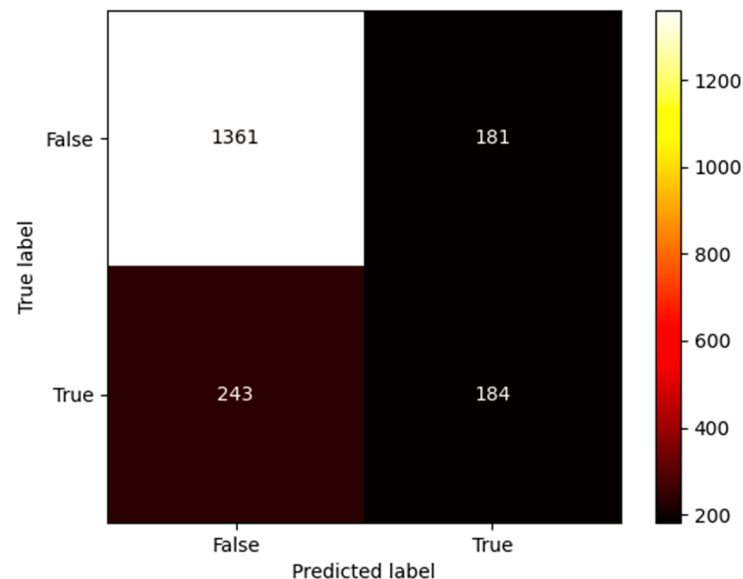


Figure 5. Pictorial representation of the confusion matrix subjected to Ethereum Fraudulence Prediction with detailed annotations. The Chaos Feature Extractor suspected 184 transactions complemented with AdaBoost to be fraudulent which were in fact fraudulent. The algorithm declared 1361 transactions to be genuine which were in fact genuine.

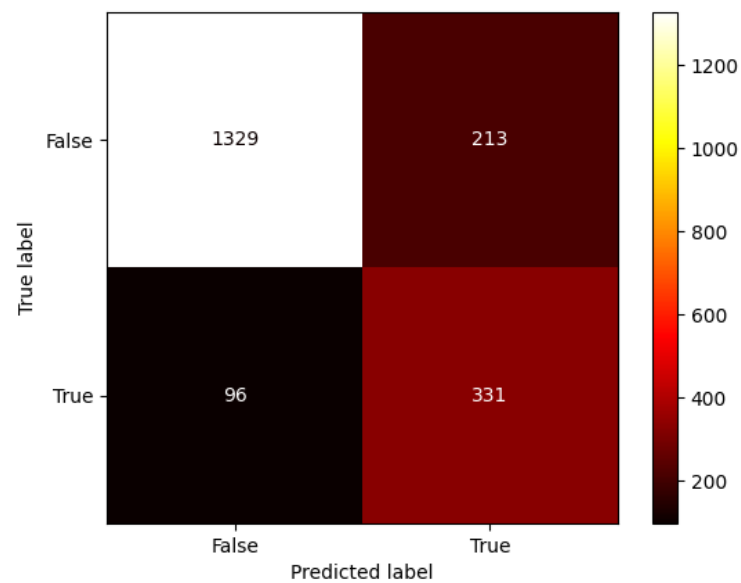


Figure 6. Pictorial representation of the confusion matrix subjected to Ethereum Fraudulence Prediction with detailed annotations. The Chaos Feature Extractor suspected 331 transactions complemented with k-nearest neighbors to be fraudulent which were in fact fraudulent. The algorithm declared 1329 transactions to be genuine which were in fact genuine.

4. Practical Implementation

For Ethereum, a decentralized framework for consensus mechanisms, Etherscan is the top blockchain analyzer, search, API, and workforce analytics tool. The creators of Etherscan created the Etherscan Developer APIs to give developers immediate access to network explorer information and features using GET/POST queries to enable universal access to the public blockchain. etherscan.db is an open-source database for keeping track of all the fraudulence in Ethereum transactions, and its open-source nature ensures that it is up-to-date. To obtain the results in Section 3 of the article, we used a chunk of 331 transactions from etherscan.db integrated with the Etherscan Developer API. For each of these, the model predicted the outcome, namely whether the transaction appeared to be fraudulent or not, and the actual result was compared with this. Based on this, the results showed that the proposed model could predict the outcome with an accuracy rate as high as 78%. Some hidden, manual sources of transactions initiated by the authors have also been tested for validation. The accuracy rate of the model was 73.9% for those transactions. The authors would like to force their readers to utilize this model to prevent them from falling victim to scams. By scams, we refer to payments made from the buyer's side in ask of some service from the lender which is then denied to them. Buyers often have to rely on trust when sending payments through cryptocurrencies since the crypto transactions are not indexed. Indeed, if buyers could make use of our model to feed in the required details described in Section 3, they could save themselves from fraudulent transactions. Further, the proposed model is not limited to Ethereum cryptocurrencies, and the same could be used for predicting fraud in other forms of cryptocurrency—namely Bitcoins, Dogecoin, etc. Any normal machine learning classifier can work out the same problem, but the benefit of using ChaosNet is that it can train the neurons of the neural network using very few data points since it is built upon the dynamics of a Chaotic Map—Generalized Luröth Series. Our world is growing at a rapid pace, but many people are not yet comfortable with carrying out transactions in cryptocurrencies. As a result of this, gathering a large amount of data is not currently possible. This could cause a problem for standard machine learning classifiers due to a lack of data.

5. Conclusions

In machine learning, making decisions when unusual events are present is difficult. This is because unusual occurrences have few data examples, which ultimately results in imbalanced learning. In this work, we have used the Neurochaos Learning (NL) architectures' usage of ChaosFEX (CFX) feature modification for imbalanced learning. Since cryptocurrencies are currently in their very nascent stage, and finding data involved in transactions in Ethereum is difficult due to their masked nature, little data can be obtained. For example, even if we examine the transaction details, they will not be sufficient to fulfill the requirements of classical ML classifier algorithms. In this work, we have tried to use ChaosNet and its indigenous Feature Extractor to try to predict possible fraud in Ethereum transactions. We used the Standalone ChaosNet, which gave us an F1 score of 0.58 for training and 0.57 for testing, which could be improved upon. Further, we used the ChaosNet Feature Extractor assisted with Adaptive Boosting to obtain an F1 score of 0.81 for training and 0.66 for testing. Finally, we used the ChaosNet Feature Extractor assisted with k-nearest neighbors to obtain an F1 score of 0.79 for training and 0.78 for testing, which was the maximum we were able to achieve. Therefore, we can conclude that the ChaosNet Feature Extractor assisted with k-nearest neighbors is the best method for predicting possible fraud in the Ethereum transaction dataset. Notably, the F1 score used here is the macro F1 score. Thus, we concluded that the final method can be used to address the issue of detecting scams in cryptocurrency transactions with an accuracy rate as high as 78%.

Future scopes of research include using the Chaos Feature Extractor in conjunction with several other ML algorithms that would result in a testing F1 score greater than 0.78. In this article, we have only tried using two well-known ML classifiers. There might be some other well-established machine learning classifiers that could improve the efficiency of the paradigm for detecting fraud in cryptocurrency transactions. Additionally, this *modus operandi* can be expanded for use with other cryptocurrencies.

Author Contributions: A.D., L.C.V. and A.R. (Athilingam Ramamoorthy)—methodology, writing, and software; L.C.V., S.R. and A.D.—data collection, writing, and editing; L.C.V., A.R. (Asif Raihan), and S.R.—writing, editing, reviewing, and methodology. A.D. and A.R. (Asif Raihan)—data visualization, draft writing, and methodology. All authors have read and agreed to the published version of the manuscript.

Funding: This research did not receive any funding.

Data Availability Statement: All data generated or analyzed during this study are publicly available here: <https://github.com/Anurag-Dutta/Ethereum> (accessed on 26 March 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Aihara, Kazuyuki, T. Takabe, and Masashi Toyoda. 1990. Chaotic neural networks. *Physics Letters A* 144: 333–40. [\[CrossRef\]](#)
- Atzei, Nicola, Massimo Bartoletti, and Tiziana Cimoli. 2017. A Survey of Attacks on Ethereum Smart Contracts (SoK). In *Principles of Security and Trust*. Lecture Notes in Computer Science. Berlin and Heidelberg: Springer, pp. 164–86. [\[CrossRef\]](#)
- Balakrishnan, Harikrishnan Nellippallil, Aditi Kathpalia, Snehanishu Saha, and Nithin Nagaraj. 2019. ChaosNet: A chaos-based artificial neural network architecture for classification. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29: 113125. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bentov, Iddo, Ariel Gabizon, and Alex Mizrahi. 2016. Cryptocurrencies without Proof of Work. In *Financial Cryptography and Data Security*. Berlin and Heidelberg: Springer, pp. 142–57. [\[CrossRef\]](#)
- Bhargavan, Karthikeyan, Antoine Delignat-Lavaud, Cédric Fournet, Anitha Gollamudi, Georges Gonthier, Nadim Kobeissi, Natalia Kulatova, Aseem Rastogi, Thomas Sibut-Pinote, Nikhil Swamy, and et al. 2016. Formal Verification of Smart Contracts. Paper presented at 2016 ACM Workshop on Programming Languages and Analysis for Security, Vienna, Austria, October 24.
- Bhavsar, Kaustubh Arun, Jimmy Singla, Yasser D. Al-Otaibi, Oh-Young Song, Yousaf Bin Zikria, and Ali Kashif Bashir. 2021. Medical Diagnosis Using Machine Learning: A Statistical Review. *Computers, Materials & Continua* 67: 107–25. [\[CrossRef\]](#)
- Chang, Hung-Jen, and Walter J. Freeman. 1996. Parameter optimization in models of the olfactory neural system. *Neural Networks* 9: 1–14. [\[CrossRef\]](#)

- Chen, Ting, Xiaoqi Li, Xiapu Luo, and Xiaosong Zhang. 2017. Under-optimized smart contracts devour your money. Paper presented at 2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER), Klagenfurt, Austria, February 20–24, pp. 442–46. [\[CrossRef\]](#)
- Choudhury, Manan Roy, and Anurag Dutta. 2022. A Perusal of Transaction Details from Silk Road 2.0 and its Cogency using the Riemann Elucidation of Integrals. *Applied Mathematics and Computational Intelligence* 11: 423–36.
- Crook, Nigel, and Tjeerd olde Scheper. 2008. Special edition of BioSystems: Information processing in cells and tissues. *Biosystems* 94: 1. [\[CrossRef\]](#)
- Decker, Christian, and Roger Wattenhofer. 2013. Information propagation in the Bitcoin network. Paper presented at IEEE P2P 2013 Proceedings, Trento, Italy, September 9–11.
- Dutta, Anurag, Manan Roy Choudhury, and Arnab Kumar De. 2022. A Unified Approach to Fraudulent Detection. *International Journal of Applied Engineering Research* 17: 110. [\[CrossRef\]](#)
- Filliâtre, Jean-Christophe, and Andrei Paskevich. 2013. Why3—Where Programs Meet Provers. In *Programming Languages and Systems*. Berlin and Heidelberg: Springer, pp. 125–28. [\[CrossRef\]](#)
- FitzHugh, Richard. 1961. Impulses and Physiological States in Theoretical Models of Nerve Membrane. *Biophysical Journal* 1: 445–66. [\[CrossRef\]](#)
- Goodell, Geoff, and Tomaso Aste. 2019. Can Cryptocurrencies Preserve Privacy and Comply with Regulations? *Frontiers in Blockchain* 2: 4. [\[CrossRef\]](#)
- Gray, Robert M. 2009. A History of Realtime Digital Speech on Packet Networks: Part II of Linear Predictive Coding and the Internet Protocol. *Foundations and Trends® in Signal Processing* 3: 203–303. [\[CrossRef\]](#)
- Harikrishnan, Nellippallil Balakrishnan, and Nithin Nagaraj. 2019. A Novel Chaos Theory Inspired Neuronal Architecture. Paper presented at 2019 Global Conference for Advancement in Technology (GCAT), Bangaluru, India, October 18–20.
- Harikrishnan, Nellippallil Balakrishnan, and Nithin Nagaraj. 2021. When Noise meets Chaos: Stochastic Resonance in Neurochaos Learning. *Neural Networks* 143: 425–35. [\[CrossRef\]](#)
- Hewamalage, Hansika, Christoph Bergmeir, and Kasun Bandara. 2021. Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. *International Journal of Forecasting* 37: 388–427. [\[CrossRef\]](#)
- Hindmarsh, James L., and R. M. Rose. 1984. A model of neuronal bursting using three coupled first order differential equations. *Proceedings of the Royal Society of London. Series B. Biological Sciences* 221: 87–102. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hodgkin, Alan L., and Andrew F. Huxley. 1952. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology* 117: 500–44. [\[CrossRef\]](#)
- Jiao, Licheng, Fan Zhang, Fang Liu, Shuyuan Yang, Lingling Li, Zhixi Feng, and Rong Qu. 2019. A Survey of Deep Learning-Based Object Detection. *IEEE Access* 7: 128837–68. [\[CrossRef\]](#)
- Jospin, Laurent Valentin, Hamid Laga, Farid Boussaid, Wray Buntine, and Mohammed Bannamoun. 2022. Hands-On Bayesian Neural Networks—A Tutorial for Deep Learning Users. *IEEE Computational Intelligence Magazine* 17: 29–48. [\[CrossRef\]](#)
- Kianpour, Mazaher, Stewart J. Kowalski, and Harald Øverby. 2021. Systematically Understanding Cybersecurity Economics: A Survey. *Sustainability* 13: 13677. [\[CrossRef\]](#)
- Kozma, Robert, and Walter J. Freeman. 1999. A possible mechanism for intermittent oscillations in the KIII model of dynamic memories—The case study of olfaction. Paper presented at IJCNN'99, International Joint Conference on Neural Networks, Proceedings (Cat. No.99CH36339), Washington, DC, USA, July 10–16. [\[CrossRef\]](#)
- Lauriola, Ivano, Alberto Lavelli, and Fabio Aioli. 2022. An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools. *Neurocomputing* 470: 443–56. [\[CrossRef\]](#)
- Makarov, Igor, and Antoinette Schoar. 2019. Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics* 135: 293–319. [\[CrossRef\]](#)
- Metcalf, William. 2020. Ethereum, Smart Contracts, DApps. *Economics, Law, and Institutions in Asia Pacific*, 77–93. [\[CrossRef\]](#)
- Meurant, Gerard. 2012. *Mass Action in the Nervous System*. Amsterdam: Elsevier.
- Montoya-Martínez, Jair, Jonas Vanthornhout, Alexander Bertrand, and Tom Francart. 2019. Effect of number and placement of EEG electrodes on measurement of neural tracking of speech. *bioRxiv*. [\[CrossRef\]](#) [\[PubMed\]](#)
- Moses, David A., Sean L. Metzger, Jessie R. Liu, Gopala K. Anumanchipalli, Joseph G. Makin, Pengfei F. Sun, Josh Chartier, Maximilian E. Dougherty, Patricia M. Liu, Gary M. Abrams, and et al. 2021. Neuroprosthesis for Decoding Speech in a Paralyzed Person with Anarthria. *New England Journal of Medicine* 385: 217–27. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nagumo, Jinichi, Suguru Arimoto, and Shuji Yoshizawa. 1962. An Active Pulse Transmission Line Simulating Nerve Axon. *Proceedings of the IRE* 50: 2061–70. [\[CrossRef\]](#)
- Sethi, Deeksha, Nithin Nagaraj, and Nellippallil Balakrishnan Harikrishnan. 2023. Neurochaos feature transformation for Machine Learning. *Integration* 90: 157–62. [\[CrossRef\]](#)
- Shen, Bo-Wen, R. A. Pielke Sr, X. Zeng, J.-J. Baik, S. Faghih-Naini, J. Cui, R. Atlas, and T. A. L. Reyes. 2021. Is Weather Chaotic? Coexisting Chaotic and Non-chaotic Attractors within Lorenz Models. In *13th Chaotic Modelling and Simulation International Conference*. Cham: Springer, pp. 805–25. [\[CrossRef\]](#)
- Sompolinsky, Yonatan, and Aviv Zohar. 2015. Secure High-Rate Transaction Processing in Bitcoin. In *Financial Cryptography and Data Security. FC 2015*. Edited by R. Böhme and T. Okamoto. Lecture Notes in Computer Science. Berlin and Heidelberg: Springer, vol. 8975. [\[CrossRef\]](#)

- Tikhomirov, Sergei. 2018. Ethereum: State of Knowledge and Research Perspectives. In *Foundations and Practice of Security*. Cham: Springer, pp. 206–21. [CrossRef]
- Trusted Smart Contracts. 2017. Available online: <http://fc17.ifca.ai/wtsc/program.html> (accessed on 19 December 2022).
- Turchin, Alexander, and Luisa F. Florez Builes. 2021. Using Natural Language Processing to Measure and Improve Quality of Diabetes Care: A Systematic Review. *Journal of Diabetes Science and Technology* 15: 553–60. [CrossRef]
- Yang, Li, and Abdallah Shami. 2020. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* 415: 295–316. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.