Influence of chaos on activation functions in Hopfield networks

Dmitriy Klyushin*,†, Oleksandr Maistrenko†

Taras Shevchenko National University of Kyiv, Volodymyrska St, 60, Kyiv, 01033, Ukraine

Abstract

Hopfield networks, renowned for their associative memory capabilities, have been extensively studied since their inception. Recent advancements in activation functions have sparked interest in exploring their applicability within the framework of these networks. This paper investigates the integration of chaotic activation functions into Hopfield networks and their implications on network dynamics and performance.

Keywords

Hopfield network, activation function, chaotic neural networks, breast cancer

1. Introduction

In deep learning, activation functions play a key role in transmitting signals between layers of neurons. While functions such as sigmoid or hyperbolic tangents have been popular in the past, recent research has emphasized the benefits of using chaos-based activation functions [1]. Chaos-based functions exhibit properties that facilitate faster and more efficient convergence during neural network training. They allow to increase the search area for optimal model parameters, helping to reduce the problem of getting stuck in local minima and providing a more stable learning process [2, 3].

This can lead to improved prediction quality and reduced time required for training deep models. Moreover, the use of chaos in activation functions can help solve the problem of training reproducibility, as these functions show less sensitivity to initial conditions [4]. However, it should be borne in mind that the use of chaos-based activation functions may require more complex optimization methods and increase computational resource requirements.

One of the mathematical models of neural network theory is the Hopfield model. A Hopfield neural network is a type of recurrent, fully connected, artificial neural network

MoMLeT-2024: 6th International Workshop on Modern Machine Learning Technologies, May, 31 - June, 1, 2024, Lviv-Shatsk, Ukraine

* Corresponding author.

† These authors contributed equally.

dmytroklyushin@knu.ua (D. Klyushin); o.maistrenko@knu.ua (O. Maistrenko)

0000-0003-4554-1049 (D. Klyushin); 0009-0003-3464-7777 (O. Maistrenko)

with a symmetric matrix of connections. In the course of operation, the dynamics of such networks converge to one of the equilibrium positions. These equilibrium positions are local minima of a functional called the network energy (in the simplest case, local minima of a negatively defined quadratic form on an n-dimensional cube). Such a network can be used as an auto-associative memory, as a filter, and to solve some optimization problems. Unlike many neural networks that work until a result is obtained after a certain number of cycles, Hopfield networks work until an equilibrium is reached, when the next state of the network is equal to the previous one. This model describes the process of training and subsequent pattern recognition.

In Hopfield models, activation functions are used to model the behavior of neurons in the brain. Although linear or sigmoidal activation functions are commonly used, some studies show that the use of chaotic activation functions can lead to interesting results [5]. For example, chaotic functions can add a stochastic element to the problem-solving process, which can be useful for some classes of problems, such as optimization or pattern recognition, where diversity and dynamics are important [6].

The use of chaotic activation functions can also improve the performance of Hopfield networks in solving problems with a large number of local minima. Chaotic functions can help in finding global optima due to their more complex and random nature [7]. But the use of chaotic activation functions can also complicate training and increase computational costs, as they require more complex optimization and processing methods.

An additional advantage of using chaotic activation functions in Hopfield networks is their ability to create more complex dynamic dependencies between neurons. This can lead to an improvement in the network's ability to adapt to changes in input data or to different environmental conditions. Such chaotic dependencies can help the model reproduce complex dynamic processes such as memory or bias learning [8, 9].

It is important to keep in mind that the use of chaotic activation functions requires additional careful tuning of model parameters and selection of appropriate training methods. Incorrect tuning can lead to unpredictable network behavior or failures in the learning process. It should also be borne in mind that chaotic activation functions may be less effective in tasks requiring high accuracy or stability, as their nature may lead to greater variability in results.

The need for a deeper study of the use of chaotic activation functions in Hopfield networks is becoming increasingly important due to the growing interest in neural networks and their potential applications in complex information processing tasks. Although some studies have already shown the promise of this approach, there is insufficient coverage of this topic in the scientific literature. It is important to conduct more in-depth experimental and theoretical research to explore the potential advantages and limitations of using chaotic activation functions in Hopfield networks [10, 11].

Understanding the dynamics of such networks with chaotic activation functions can lead to important discoveries in the areas of information storage and recovery, as well as in working with large data sets.

2. Chaos in Hopfield networks

Let the variable x(t) denote the activity of the neuron at time t. The state x(t) = 1 corresponds to the "excitation" state, and the state x(t) = -1 corresponds to the "inhibition" state. Let h(t) be an external influence on the neuron, for example, from other neurons. For convenience, we assume that time is discrete (t = 0, 1, 2,...). In [12], it was proposed that the evolution of a neuron is determined by a dynamical system, which is a piecewise linear, piecewise continuous one-dimensional function:

$$x(t + l) = F(x(t), h(t)) \quad (1)$$

This family of functions is defined by the function K(h) of the slope angle. This dependence must satisfy the following conditions: 1) parity K(-h) = K(h); 2) monotonic decrease at h > 0 (and, therefore, monotonic increase at h < 0); 3) K(0) = 2. All these requirements are met for the function

$$K(h) = \frac{2}{1 + |h|/\mu} \quad (2)$$

The dynamics of a formal neuron is determined by a ratio:

$$x(t+1) = \begin{cases} 1, & h(t) > 0 \\ -1, & h(t) < 0 \end{cases} \quad (3)$$

That is, without any transient process, x(t) takes the value +1 or - 1 regardless of the neuron's state at the previous moment.

It is known that a real neuron is an extremely complex system of inertia [13, 14, 15, 16]. Its processes have characteristic times ranging from units to hundreds (or more) of milliseconds [17, 18, 19, 20]. In [12], it was assumed that at a large value of external influence h, the neuron "quickly" enters the state of excitation, while at lower values, the speed of transition to the state of excitation decreases. This is taken into account in the above chaotic Izhikevich activation function in (1), because as h decreases, K(h) increases and the transition time increases in (2), although the only special point x = 1 is still stable. When K(h) passes through 1, x = 1 loses stability and the neuron's dynamics becomes chaotic.

The above scenario of the transition from order to chaos in (3) is observed in many discontinuous or sharp-vertex functions and is called a homoclinic explosion [21]. Previous studies have shown that the results are weakly dependent on the specific type of the function K(h) if this scenario is chaos. The most important parameter to display is only the value of h = p, in which K(h) = 1, i.e., at which the transition to chaos occurs.

3. Comparison of activation functions

Differentiable activation functions in neural networks have a number of significant advantages over their undifferentiable counterparts. First of all, they allow efficient use of optimization algorithms, such as back-propagation, to train the network. The smooth nature of differentiable functions allows you to find optimal parameters by searching for gradients, which simplifies the training process and improves the convergence rate.

It is important to keep in mind that differentiable activation functions allow the use of regularization methods that reduce network overtraining and improve its overall generalization ability. For example, you can apply L1 or L2 regularization to the model weights, which helps to control their complexity and avoid overfitting.

In addition, differentiable activation functions provide a smooth and continuous relationship between the network's inputs and outputs, which facilitates smooth transitions between values and eases optimization processes. They also allow the use of more powerful optimization algorithms, such as adaptive gradient descent methods, which contributes to faster and more efficient network convergence.

The chaotic Izhikevich activation function described in the previous section has a drawback, it is not differentiable for small values of h(t). This drawback can be solved by replacing the linear function with a parametric Bézier function (check Figure 1). Such a replacement allows to ensure the differentiability of activation functions by fulfilling the condition on the bijectionality of the equation x(p), where x is a vector of input values and p is the parameters of the Bézier function.

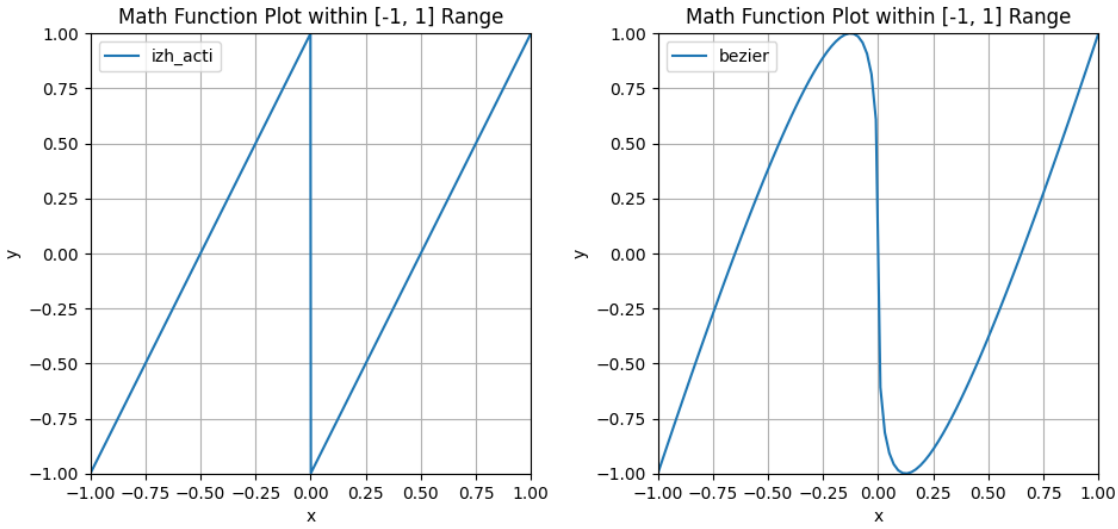

Figure 1: The chaotic Izhikevich activation function and its modified version with a parametric Bezier curve

The following experiment was conducted to compare the activation functions of the Hopfield network - a signum, a chaotic Izhikevich activation function, and its differentiable modification. The input data were samples from normal distributions $N(\mu, \sigma^2)$ of two classes. The parameters μ and σ^2 were chosen to cover a variety of cases of intersection of the distributions. Training images were built on the training data according to each category of binary classification, and the quality of the models was evaluated on the test data using Leave-One-Out validation, with the F1-score as the target metric.

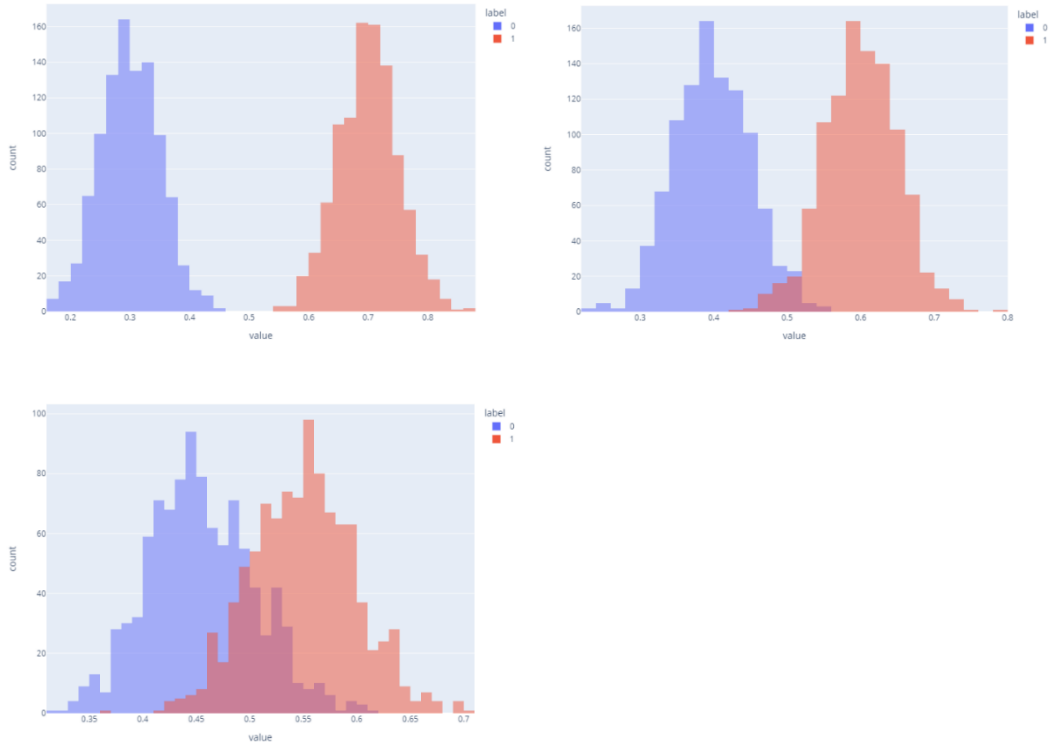

Figure 2: Samples for binary classification from normal distributions with the same variance and different mathematical expectation

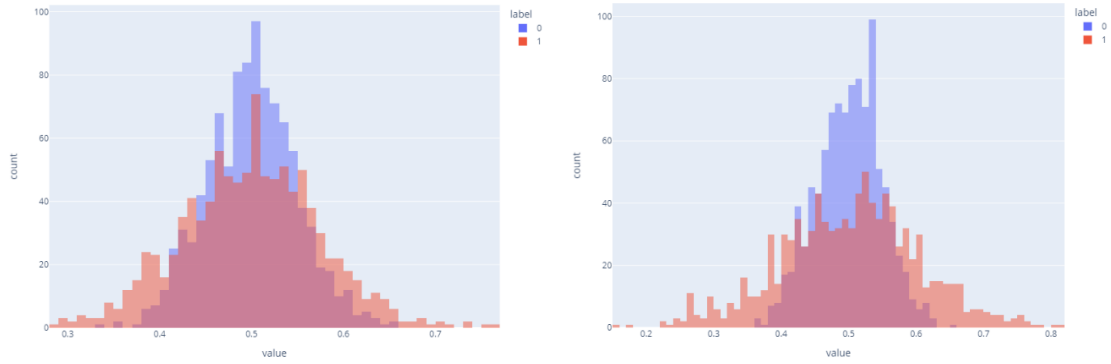

Figure 3: Samples for binary classification from normal distributions with the same mathematical expectation and different variance

The null hypothesis is that the activation functions do not affect the model results. The alternative hypothesis is that chaotic activation functions produce better results.

Consider the case when the distributions intersect, $\mu_1 = 0.4, \mu_2 = 0.6, \sigma_1^2 = \sigma_2^2 = 0.05$. After 100 experiments, using the signum as an activation function, an average F1-score of

0.81 was obtained. Using statistical power, the required number of experiments for the incremental increase to the mean value of the F1-score to be significant can be obtained. Accordingly, at least 23664 experiments are needed for each group for the z-test to be significant for an increase of at least 1%.

24000 experiments were conducted for each of the activation functions compared (signum, chaotic Izhikevich activation function and its differentiable modification), and calculated F1-scores for each experiment. The average value of the F1-score for the experiments with the signum was 0.9, with the chaotic Izhikevich activation function was 0.91, and with the differentiable modification was 0.911. After that, z-tests were performed for each pair of experiments (check Table 1).

Table 1
Results of z-tests for experiments with activation functions

N	Title	z-statistics	p-value
1	sign_vs_izh	-4.8446	< 0.0001
2	sign_vs_bezier	-5.022	< 0.0001
3	izh_vs_bezier	-1.5722	0.1158

Both samples with chaotic functions are statistically significantly different from the sample of experiments with signum, but do not have a statistically significant difference between themselves, although the modified function gave better results, as can be seen from the z-statistics.

Similarly, experiments with other parameters of normal distributions were conducted, calculated F1-scores for each sample, and conducted paired z-tests (check Table 2).

Table 2
Results of z-tests for experiments with activation functions

N	Title	Parameters	z-statistics	p-value
1	sign_vs_izh	$\mu_1 = 0.3, \mu_2 = 0.7, \sigma_1^2 = \sigma_2^2 = 0.05$	-5.6611	< 0.0001
2	sign_vs_bezier	$\mu_1 = 0.3, \mu_2 = 0.7, \sigma_1^2 = \sigma_2^2 = 0.05$	-5.4636	< 0.0001
3	izh_vs_bezier	$\mu_1 = 0.3, \mu_2 = 0.7, \sigma_1^2 = \sigma_2^2 = 0.05$	-1.4194	0.1557
4	sign_vs_izh	$\mu_1 = 0.45, \mu_2 = 0.55, \sigma_1^2 = \sigma_2^2 = 0.05$	-3.0375	0.0023
5	sign_vs_bezier	$\mu_1 = 0.45, \mu_2 = 0.55, \sigma_1^2 = \sigma_2^2 = 0.05$	-3.4627	0.0005
6	izh_vs_bezier	$\mu_1 = 0.45, \mu_2 = 0.55, \sigma_1^2 = \sigma_2^2 = 0.05$	-0.3879	0.6980
7	sign_vs_izh	$\mu_1 = \mu_2 = 0.5, \sigma_1^2 = 0.05, \sigma_2^2 = 0.1$	-1.6876	0.0914
8	sign_vs_bezier	$\mu_1 = \mu_2 = 0.5, \sigma_1^2 = 0.05, \sigma_2^2 = 0.1$	-1.3331	0.1824
9	izh_vs_bezier	$\mu_1 = \mu_2 = 0.5, \sigma_1^2 = 0.05, \sigma_2^2 = 0.1$	0.9344	0.3500
10	sign_vs_izh	$\mu_1 = \mu_2 = 0.5, \sigma_1^2 = 0.05, \sigma_2^2 = 0.075$	-1.2260	0.2199
11	sign_vs_bezier	$\mu_1 = \mu_2 = 0.5, \sigma_1^2 = 0.05, \sigma_2^2 = 0.075$	-1.5949	0.1107
12	izh_vs_bezier	$\mu_1 = \mu_2 = 0.5, \sigma_1^2 = 0.05, \sigma_2^2 = 0.075$	-0.4813	0.6302

For the experiments with different mathematical expectation, the previous tendency has been preserved: chaotic functions show a statistically significant improvement over signum, but the differentiable modification shows better results.

In the experiments with the same mathematical expectation, all functions showed similar results and have no statistically significant difference.

4. Experiment on fractal analysis of buccal epithelium kernels

Now let's consider the effectiveness of chaotic activation functions in an experiment with real data.

Some of the first reports of malignant changes appeared in the 1960s, when the content of X-chromatin in somatic cells was widely studied and its instability became apparent in various functional changes in the body and general pathology of somatic cells. In the presence of tumors in the body, there are obvious changes in the content of X-chromatin in the buccal epithelium and peripheral blood neutrophils. It has also been shown that changes in the number of cells with X chromosomes are caused by a defect in the functional state of the heterozygous X-chromosome [22].

Studies demonstrating changes in buccal epithelial cells in patients with tumors are of great interest: in the 1960s, H. Nieburgs and co-authors [23] reported a characteristic redistribution of chromatin mass in somatic cells in 77% of cancer patients and called these changes tumorigenic. These changes were characterized by an increase in the size of epithelial cell nuclei and an increase in the size of "restricted" chromatin regions surrounded by light areas [24]. The same changes were observed in cells of the liver, kidneys and other organs. In patients with breast cancer, an increase in DNA content in the buccal epithelium and the size of interphase nuclei was found. However, some authors did not find a significant difference in this indicator between patients and practically healthy men when the DNA content of buccal epithelial cells in men with bronchial epithelioma was measured by cellular spectrophotometry [25].

Study [26] examined a control group (29 patients), a group of patients with stage II breast cancer (68 patients), and a group of patients with fibroadenomatosis (33 patients). All diagnoses were confirmed by histology. The morphologic dataset consisted of 20256 images of interphase nuclei of the buccal epithelium (6752 nuclei were scanned in three versions: without filtering, with a yellow filter, and with a purple filter).

The morphological material was smears of epithelial cells of the oral mucosa in the middle depth of the spinous layer. On average, each preparation consisted of 52 cells. The content of DNA-fuchsin in the nuclei of epithelial cells was calculated as the product of optical density and area. At the first stage of the analysis, an image of the chromatin distribution was obtained in the form of a 128 x 128 pixel matrix [26].

To reflect the fractal nature of the chromatin distribution and to ensure invariance to image rotation, a spatial fill curve was created along each pixel of the image, and the RGB color values of the image pixels were read sequentially rather than line by line. As a result, the pixel matrix could be mapped to three vectors corresponding to the three channels of the RGB color model. The Sierpinski curves [26] were used as space filling curves.

In order to apply the fractal image analysis method, the image must be pre-processed. For this purpose, the Otsu method was applied [26]. This method is used for threshold binarization of halftone images. The algorithm assumes that there are two classes of pixels

in the image (main and background), and searches for the optimal threshold value that divides the pixels into two classes so that the intra-class variance is minimal.

There are several methods for calculating the fractal dimension of an image. In [26], the Hurst index was chosen because it is very suitable for sequential analysis. The Hurst index is related to the fractal dimension D by the formula H = 2 - D.

The input data for the model were three-channel (RGB) samples of fractal kernel dimensions for each patient. The samples differed significantly in the number of elements. Therefore, when preparing the data before training the neural network, n quantiles were calculated for each sample, where n is the number of elements in the smallest sample (check Figure 4).

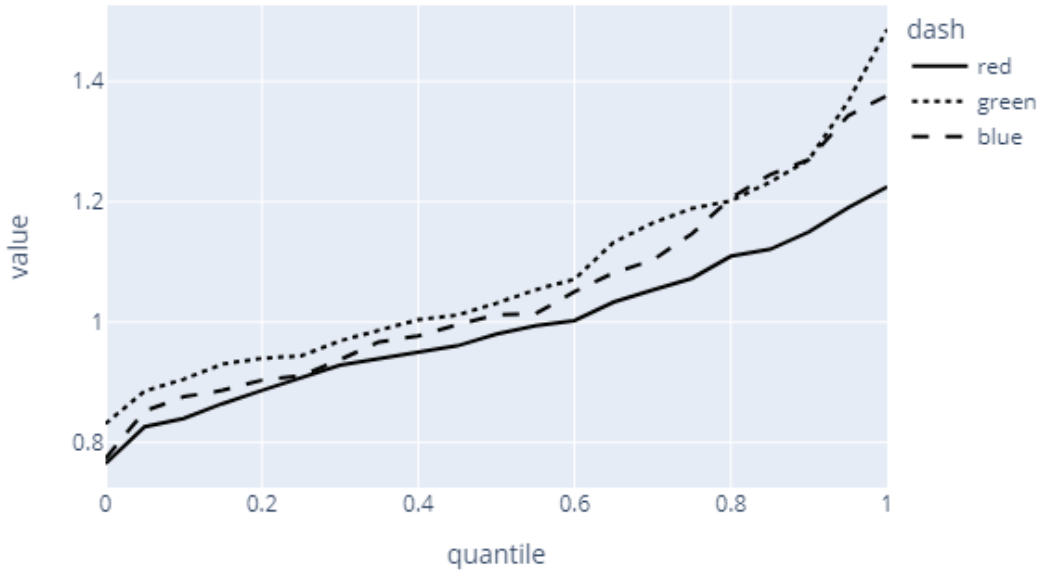

Figure 4: An example of consolidation of training samples using quantiles

Additional datasets were created from the input three-channel RGB data. The following input data samples were used in the experiments: RGB (three channels at the same time), R (red only), G (green only), B (blue only), Gray (gray channel calculated using the formula $0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B$), Mean (arithmetic mean of three channels).

To build training images for the Hopfield network, two methods of quantile aggregation were used: arithmetic mean and median.

At this stage, the training images were sequences of real numbers. The neurons of the Hopfield network can have two values: 1 or -1. Therefore, additional binarization was applied to the images: real values were rounded to the ninth decimal place, converted to binary and concatenated into the final binary portrait (check Figure 5).

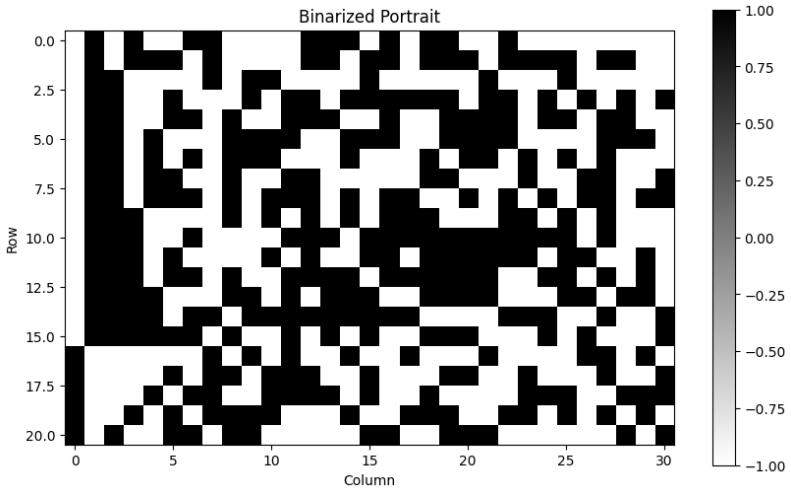

Figure 5: An example of prepared training image

A total of 144 experiments were set up and divided into 4 categories: cancer patients vs. healthy controls, cancer patients vs. healthy controls and fibroadenomatosis patients, cancer patients vs. fibroadenomatosis patients, fibroadenomatosis patients vs. healthy controls.

Training images were built on the training data according to each category of binary classification, the quality of the models was evaluated on the test data using Leave-One-Out validation, and the target metrics were precision, recall, and F1-score.

Table 3

Results of the comparative experiment on fractal analysis of
buccal epithelium kernels (cancer patients vs. fibroadenomatosis patients)

N	Experiment	Input	Aggregation	F. activation	Precision	Recall	F1
1	BC_vs_FAM	B	avg	bezier	0.6389	0.3382	0.4423
2	BC_vs_FAM	B	avg	izh	0.6316	0.3529	0.4528
3	BC_vs_FAM	B	avg	sign	0.6750	0.3971	0.5000
4	BC_vs_FAM	B	median	bezier	0.8077	0.6176	0.7000
5	BC_vs_FAM	B	median	izh	0.8269	0.6324	0.7167
6	BC_vs_FAM	B	median	sign	0.8077	0.6176	0.7000
7	BC_vs_FAM	G	avg	bezier	0.6538	0.5000	0.5667
8	BC_vs_FAM	G	avg	izh	0.6400	0.4706	0.5424
9	BC_vs_FAM	G	avg	sign	0.6538	0.5000	0.5667
10	BC_vs_FAM	G	median	bezier	0.8727	0.7059	0.7805
11	BC_vs_FAM	G	median	izh	0.8679	0.6765	0.7603
12	BC_vs_FAM	G	median	sign	0.8727	0.7059	0.7805
13	BC_vs_FAM	gray	avg	bezier	0.7292	0.5147	0.6034
14	BC_vs_FAM	gray	avg	izh	0.7347	0.5294	0.6154
15	BC_vs_FAM	gray	avg	sign	0.7292	0.5147	0.6034
16	BC_vs_FAM	gray	median	bezier	0.7347	0.5294	0.6154
17	BC_vs_FAM	gray	median	izh	0.7059	0.5294	0.6050
18	BC_vs_FAM	gray	median	sign	0.7347	0.5294	0.6154

19	BC_vs_FAM	mean	avg	bezier	0.6667	0.4706	0.5517
20	BC_vs_FAM	mean	avg	izh	0.6600	0.4853	0.5593
21	BC_vs_FAM	mean	avg	sign	0.6735	0.4853	0.5641
22	BC_vs_FAM	mean	median	bezier	0.7143	0.5882	0.6452
23	BC_vs_FAM	mean	median	izh	0.7193	0.6029	0.6560
24	BC_vs_FAM	mean	median	sign	0.7143	0.5882	0.6452
25	BC_vs_FAM	R	avg	bezier	0.6800	0.5000	0.5763
26	BC_vs_FAM	R	avg	izh	0.6596	0.4559	0.5391
27	BC_vs_FAM	R	avg	sign	0.5429	0.2794	0.3689
28	BC_vs_FAM	R	median	bezier	0.8636	0.8382	0.8507
29	BC_vs_FAM	R	median	izh	0.8636	0.8382	0.8507
30	BC_vs_FAM	R	median	sign	0.8636	0.8382	0.8507
31	BC_vs_FAM	RGB	avg	bezier	0.7333	0.4853	0.5841
32	BC_vs_FAM	RGB	avg	izh	0.6923	0.3971	0.5047
33	BC_vs_FAM	RGB	avg	sign	0.6667	0.3529	0.4615
34	BC_vs_FAM	RGB	median	bezier	0.9474	0.7941	0.8640
35	BC_vs_FAM	RGB	median	izh	0.9310	0.7941	0.8571
36	BC_vs_FAM	RGB	median	sign	0.9474	0.7941	0.8640

In the comparison "cancer patients vs. fibroadenomatosis patients", the differentiable modification showed better results in the experiments with the best results, the median being a better aggregation function than the arithmetic mean. The highest accuracy was obtained with input data in RGB format and single-channel inputs R, G, and B (check Table 3).

Table 4
Results of the comparative experiment on fractal analysis of buccal epithelium kernels (cancer patients vs. healthy controls)

N	Experiment	Input	Aggregation	F. activation	Precision	Recall	F1
1	BC_vs_C	B	avg	bezier	0.8980	0.6471	0.7521
2	BC_vs_C	B	avg	izh	0.9000	0.6618	0.7627
3	BC_vs_C	B	avg	sign	0.8980	0.6471	0.7521
4	BC_vs_C	B	median	bezier	0.8364	0.6765	0.7480
5	BC_vs_C	B	median	izh	0.8364	0.6765	0.7480
6	BC_vs_C	B	median	sign	0.8364	0.6765	0.7480
7	BC_vs_C	G	avg	bezier	0.6591	0.4265	0.5179
8	BC_vs_C	G	avg	izh	0.6744	0.4265	0.5225
9	BC_vs_C	G	avg	sign	0.6591	0.4265	0.5179
10	BC_vs_C	G	median	bezier	0.7347	0.5294	0.6154
11	BC_vs_C	G	median	izh	0.7174	0.4853	0.5789
12	BC_vs_C	G	median	sign	0.7347	0.5294	0.6154
13	BC_vs_C	gray	avg	bezier	0.7500	0.5735	0.6500
14	BC_vs_C	gray	avg	izh	0.7692	0.5882	0.6667
15	BC_vs_C	gray	avg	sign	0.7500	0.5735	0.6500
16	BC_vs_C	gray	median	bezier	0.7059	0.5294	0.6050
17	BC_vs_C	gray	median	izh	0.7000	0.5147	0.5932
18	BC_vs_C	gray	median	sign	0.7059	0.5294	0.6050

19	BC_vs_C	mean	avg	bezier	0.7556	0.5000	0.6018
20	BC_vs_C	mean	avg	izh	0.7660	0.5294	0.6261
21	BC_vs_C	mean	avg	sign	0.7556	0.5000	0.6018
22	BC_vs_C	mean	median	bezier	0.7917	0.5588	0.6552
23	BC_vs_C	mean	median	izh	0.7917	0.5588	0.6552
24	BC_vs_C	mean	median	sign	0.7917	0.5588	0.6552
25	BC_vs_C	R	avg	bezier	0.7347	0.5294	0.6154
26	BC_vs_C	R	avg	izh	0.7347	0.5294	0.6154
27	BC_vs_C	R	avg	sign	0.7347	0.5294	0.6154
28	BC_vs_C	R	median	bezier	0.7143	0.5147	0.5983
29	BC_vs_C	R	median	izh	0.7143	0.5147	0.5983
30	BC_vs_C	R	median	sign	0.7143	0.5147	0.5983
31	BC_vs_C	RGB	avg	bezier	0.8889	0.5882	0.7080
32	BC_vs_C	RGB	avg	izh	0.8636	0.5588	0.6786
33	BC_vs_C	RGB	avg	sign	0.8696	0.5882	0.7018
34	BC_vs_C	RGB	median	bezier	0.7959	0.5735	0.6667
35	BC_vs_C	RGB	median	izh	0.7959	0.5735	0.6667
36	BC_vs_C	RGB	median	sign	0.7959	0.5735	0.6667

In the comparison "cancer patients vs. healthy controls", the differentiable modification showed similar results to the chaotic Izhikevich function. Both chaotic functions gave better results than signum in most experiments. The median and arithmetic mean gave similar results in similar experiments. The most accurate results were obtained with input data in B and RGB formats (check Table 4).

Table 5

Results of the comparative experiment on fractal analysis of buccal epithelium kernels (cancer patients vs. healthy controls and fibroadenomatosis patients)

N	Experiment	Input	Aggregation	F. activation	Precision	Recall	F1
1	BC_vs_C+FAM	B	avg	bezier	0.6207	0.5294	0.5714
2	BC_vs_C+FAM	B	avg	izh	0.6000	0.5294	0.5625
3	BC_vs_C+FAM	B	avg	sign	0.6316	0.5294	0.5760
4	BC_vs_C+FAM	B	median	bezier	0.5857	0.6029	0.5942
5	BC_vs_C+FAM	B	median	izh	0.5857	0.6029	0.5942
6	BC_vs_C+FAM	B	median	sign	0.5857	0.6029	0.5942
7	BC_vs_C+FAM	G	avg	bezier	0.5507	0.5588	0.5547
8	BC_vs_C+FAM	G	avg	izh	0.5606	0.5441	0.5522
9	BC_vs_C+FAM	G	avg	sign	0.5507	0.5588	0.5547
10	BC_vs_C+FAM	G	median	bezier	0.5254	0.4559	0.4882
11	BC_vs_C+FAM	G	median	izh	0.5345	0.4559	0.4921
12	BC_vs_C+FAM	G	median	sign	0.5254	0.4559	0.4882
13	BC_vs_C+FAM	gray	avg	bezier	0.5588	0.5588	0.5588
14	BC_vs_C+FAM	gray	avg	izh	0.5571	0.5735	0.5652
15	BC_vs_C+FAM	gray	avg	sign	0.5588	0.5588	0.5588
16	BC_vs_C+FAM	gray	median	bezier	0.6034	0.5147	0.5556
17	BC_vs_C+FAM	gray	median	izh	0.6034	0.5147	0.5556

18	BC_vs_C+FAM	gray	median	sign	0.6034	0.5147	0.5556
19	BC_vs_C+FAM	mean	avg	bezier	0.5333	0.4706	0.5000
20	BC_vs_C+FAM	mean	avg	izh	0.5469	0.5147	0.5303
21	BC_vs_C+FAM	mean	avg	sign	0.5333	0.4706	0.5000
22	BC_vs_C+FAM	mean	median	bezier	0.6071	0.5000	0.5484
23	BC_vs_C+FAM	mean	median	izh	0.6140	0.5147	0.5600
24	BC_vs_C+FAM	mean	median	sign	0.6071	0.5000	0.5484
25	BC_vs_C+FAM	R	avg	bezier	0.5294	0.5294	0.5294
26	BC_vs_C+FAM	R	avg	izh	0.5441	0.5441	0.5441
27	BC_vs_C+FAM	R	avg	sign	0.5294	0.5294	0.5294
28	BC_vs_C+FAM	R	median	bezier	0.5741	0.4559	0.5082
29	BC_vs_C+FAM	R	median	izh	0.5741	0.4559	0.5082
30	BC_vs_C+FAM	R	median	sign	0.5741	0.4559	0.5082
31	BC_vs_C+FAM	RGB	avg	bezier	0.5667	0.5000	0.5313
32	BC_vs_C+FAM	RGB	avg	izh	0.5738	0.5147	0.5426
33	BC_vs_C+FAM	RGB	avg	sign	0.5667	0.5000	0.5313
34	BC_vs_C+FAM	RGB	median	bezier	0.6000	0.5294	0.5625
35	BC_vs_C+FAM	RGB	median	izh	0.6066	0.5441	0.5736
36	BC_vs_C+FAM	RGB	median	sign	0.6000	0.5294	0.5625

In the comparison "cancer patients vs. healthy controls and fibroadenomatosis patients", all three activation functions showed similar results. Similarly, there are no overperforming alternatives among the aggregations and input data formats (check Table 5).

Table 6

Results of the comparative experiment on fractal analysis of buccal epithelium kernels (fibroadenomatosis patients vs. healthy controls)

N	Experiment	Input	Aggregation	F. activation	Precision	Recall	F1
1	FAM_vs_C	B	avg	bezier	0.6429	0.5625	0.6000
2	FAM_vs_C	B	avg	izh	0.5926	0.5000	0.5424
3	FAM_vs_C	B	avg	sign	0.6429	0.5625	0.6000
4	FAM_vs_C	B	median	bezier	0.7353	0.7813	0.7576
5	FAM_vs_C	B	median	izh	0.7429	0.8125	0.7761
6	FAM_vs_C	B	median	sign	0.7353	0.7813	0.7576
7	FAM_vs_C	G	avg	bezier	0.4583	0.3438	0.3929
8	FAM_vs_C	G	avg	izh	0.4231	0.3438	0.3793
9	FAM_vs_C	G	avg	sign	0.4400	0.3438	0.3860
10	FAM_vs_C	G	median	bezier	0.4688	0.4688	0.4688
11	FAM_vs_C	G	median	izh	0.4516	0.4375	0.4444
12	FAM_vs_C	G	median	sign	0.4688	0.4688	0.4688
13	FAM_vs_C	gray	avg	bezier	0.4800	0.3750	0.4211
14	FAM_vs_C	gray	avg	izh	0.4815	0.4063	0.4407
15	FAM_vs_C	gray	avg	sign	0.4615	0.3750	0.4138
16	FAM_vs_C	gray	median	bezier	0.4688	0.4688	0.4688
17	FAM_vs_C	gray	median	izh	0.4333	0.4063	0.4194
18	FAM_vs_C	gray	median	sign	0.4688	0.4688	0.4688

19	FAM_vs_C	mean	avg	bezier	0.6429	0.5625	0.6000
20	FAM_vs_C	mean	avg	izh	0.6552	0.5938	0.6230
21	FAM_vs_C	mean	avg	sign	0.6000	0.5625	0.5806
22	FAM_vs_C	mean	median	bezier	0.4444	0.5000	0.4706
23	FAM_vs_C	mean	median	izh	0.4444	0.5000	0.4706
24	FAM_vs_C	mean	median	sign	0.4444	0.5000	0.4706
25	FAM_vs_C	R	avg	bezier	0.6667	0.5625	0.6102
26	FAM_vs_C	R	avg	izh	0.6000	0.5625	0.5806
27	FAM_vs_C	R	avg	sign	0.5000	0.5625	0.5294
28	FAM_vs_C	R	median	bezier	0.5200	0.4063	0.4561
29	FAM_vs_C	R	median	izh	0.4815	0.4063	0.4407
30	FAM_vs_C	R	median	sign	0.5200	0.4063	0.4561
31	FAM_vs_C	RGB	avg	bezier	0.7083	0.5313	0.6071
32	FAM_vs_C	RGB	avg	izh	0.6538	0.5313	0.5862
33	FAM_vs_C	RGB	avg	sign	0.5862	0.5313	0.5574
34	FAM_vs_C	RGB	median	bezier	0.6774	0.6563	0.6667
35	FAM_vs_C	RGB	median	izh	0.7000	0.6563	0.6774
36	FAM_vs_C	RGB	median	sign	0.6774	0.6563	0.6667

In the comparison "fibroadenomatosis patients vs. healthy controls", chaotic functions showed better results than signum in most experiments. The median was more effective than the arithmetic mean in similar experiments. Input data in RGB and B formats gave the highest accuracy (check Table 6).

The best results were obtained in the experiments "cancer patients vs. fibroadenomatosis patients" - the average value of the F1-score of the 5 best experiments is 0.84, the worst results were obtained in the group of experiments "cancer patients vs. healthy controls and fibroadenomatosis patients " - the average value of the F1-score of the 5 best experiments is 0.57. This can be explained by the distinction of the second class – healthy controls and fibroadenomatosis patients - which, when aggregated, generated an image that the model could hardly distinguish from cancer patients.

Also, the median proved to be a better aggregation for training images - the F1-score is higher in most experiments.

The best input sample formats are RGB and B, which are present in the top five for each group of experiments.

5. Conclusions

In deep learning, especially in the context of artificial neural networks, the use of chaos in activation functions plays a significant role in ensuring the efficiency and flexibility of models. Chaotic activation functions allow to enrich the feature distribution space, which allows the model to adapt to various conditions and maintain resistance to noise in the data. This approach increases the robustness of models to changes in training data and can improve their overall ability to generalize to new data. The use of chaotic activation functions can also be important in cases where the model needs to adapt to unpredictable or nonlinear relationships in the data, making it more versatile and powerful in solving complex machine learning problems.

Also, activation functions that can be differentiated in neural networks have several important advantages over their undifferentiable counterparts. First of all, they allow for the efficient use of optimization algorithms, such as back-propagation of errors, to train the network. The smooth nature of differentiable functions allows finding optimal parameters using gradient descent, which simplifies the training process and improves the convergence rate.

The paper substantiates and demonstrates the advantages of chaotic activation functions of the Hopfield model on the example of experiments with artificial samples from normal distributions and on real data in predicting breast cancer.

This area still requires further research and experimentation, and the development and implementation of new chaotic activation functions may be a promising way to improve deep learning not only in Hopfield models but also in alternative algorithms and neural network architectures and improve results in various tasks.

Acknowledgements

Authors would like to express their sincere gratitude to K. Golubeva and N. Borodai for generously providing the data used for training the models in this study.

References

[1] S. R. Dubey, S. K. Singh, B. B. Chaudhuri, Activation Functions in Deep Learning: A comprehensive Survey and Benchmark, Neurocomputing 503 (2022) 24-27. doi:10.1016/j.neucom.2022.06.111.

[2] S. Kilicarslan, K. Adem, M. Celik, An overview of the activation functions used in deep learning algorithms, Journal of New Results in Science 10 (2021) 75-88. doi:10.54187/jnrs.1011739.

[3] X. Wang, H. Ren, A. Wang, Smish: A Novel Activation Function for Deep Learning Methods, Electronics 11 (2022) 540-545. doi:10.3390/electronics11040540.

[4] X. Chen, Y. Wang, A Chaotic Neuron and its Ability to Prevent Overfitting, Frontiers in Computing and Intelligent Systems 5 (2023) 53-61. doi:10.54097/fcis.v5i1.11673.

[5] Q. Deng, C. Wong, H. Lin, Memristive Hopfield neural network dynamics with heterogeneous activation functions and its application, Chaos Solitons & Fractals 178 (2024) 101-124. doi:10.1016/j.chaos.2023.114387.

[6] Q. Deng, C. Wong, H. Lin, Chaotic dynamical system of Hopfield neural network influenced by neuron activation threshold and its image encryption, Nonlinear Dynamics 112 (2024) 1-18. doi:10.1007/s11071-024-09384-3.

[7] D. Magallon, J. Garcia, G. Huerta, R. Jaimes, Real-time neural identification using a recurrent wavelet first-order neural network of a chaotic system implemented in an FPAA, Integration 96 (2024) 384-393. doi:10.1016/j.vlsi.2023.102134.

[8] D. Krotov, A new frontier for Hopfield networks, Nature Reviews Physics 5 (2023) 52-70. doi:10.1038/s42254-023-00595-y.

[9] C. Cursino, L. Dias, Hybrid Hopfield Neural Network, SN Computer Science 5 (2024) 25-74. doi:10.1007/s42979-023-02575-6.

[10] S. Kashyap, N. Dsouza, L. Shi, D. Beymer, Hopfield Encoding Networks (2022) doi:10.13140/RG.2.2.17209.65123.

[11] X. Liang, Y. Yang, R. Wang, J. Chen, Synchronization of delayed stochastic reaction-diffusion Hopfield neural networks via sliding mode control, Nonlinear Analysis Modelling and Control 29 (2024) 1-19. doi:10.15388/namc.2024.29.34884.
[12] E. Lieberman-Aiden, Comprehensive mapping of long-range interactions reveals folding principles of the human Genome, Science 326 (2009) 289-193. doi:10.1126/science.1181369.
[13] A. Babloyantz, S. Nicolis, M. Salazor, Evidence of chaotic dynamics of brain activity during the sleep cycle, Physics Letters A. 111 (1985) 152-156. doi:10.1016/0375-9601(85)90444-X.
[14] W. Freeman, Simulation of Chaotic EEG Patterns with a Dynamic Model of the Olfactory System, Biological Cybernetics 56 (1987) 139-150. doi:10.1007/BF00317988.
[15] M. Hirsch, Convergent Activation Dynamics in Continuous Time Neural Networks, Neural Networks 2 (1989) 331-351. doi:10.1016/0893-6080(89)90018-X.
[16] J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc Natl Acad Sci U S A. Apr; 79 (1982) 2554-2558. doi:10.1073/pnas.79.8.2554.
[17] G. Shepherd, The synaptic organization of the brain, Oxford University Press (1990) doi:10.1093/acprof:oso/9780195159561.001.1.
[18] C. Skarda, W. Freeman, How brains make chaos in order to make sense of the world, Behavioral and Brain Sciences 10 (1987) 161-195. doi:10.1017/S0140525X00047336.
[19] I. Tsuda, Dynamic link of memory - Chaotic memory map in nonequilibrium neural networks, Neural Networks 5 (1992) 857-893. doi:10.1016/S0893-6080(05)80029-2.
[20] Y. Yao, W. Freeman, Model of Biological Pattern Recognition with Spatially Chaotic Dynamics, Neural Networks 3 (1990) 153 - 170. doi:10.1016/0893-6080(90)90086-Z.
[21] S. Sparrow, The Lorenz Equations. Bifurcations, Chaos, and Strange Attractors, Applied Mathematical Sciences 41 (1982) 26-50. doi:10.1007/978-1-4612-5767-7.
[22] M. Radak, H. Lafta, H. Fallahi, Machine learning and deep learning techniques for breast cancer diagnosis and classification: a comprehensive review of medical imaging studies, Journal of Cancer Research and Clinical Oncology 49 (2023) 10473-10491. doi:10.1007/s00432-023-04956-z.
[23] H. Nieburgs, Recent progress in the interpretation of malignancy associated changes (MAC), Acta Cytologica 12 (1968) 445-453. doi:10.1155/238921.
[24] N. Boroday, V. Chekhun, E. Golubeva, D. Klyushin, In vitro and in vivo densitometric analysis of DNA content and chromatin texture in tumor cell nuclei under the influence of a nano composite and magnetic field, Advances in Cancer Research & Treatment 2016 (2016) 1-11. doi:10.5171/2016.706183.
[25] G. Ogden, J. Cowpe, M. Green, The effect of distant malignancy upon quantitative cytologic assessment of normal oral mucosa, Cancer 65 (1990) 477-480. doi:10.1002/1097-0142(19900201)65:3<477::AID-CNCR2820650317>3.0.CO;2-G.
[26] D. Klyushin, K. Golubeva, N. Boroday, D. Shervarly, Breast Cancer Diagnosis Using Machine Learning and Fractal Analysis of Malignancy-Associated Changes in Buccal Epithelium, Artificial Intelligence, Machine Learning, and Data Science Technologies Future Impact and Well- Being for Society 5 (2021) 1-18. doi:10.1201/9781003153405-1.