Web Phishing Detection System Based on Artificial Neural Networks Technology

Vitaliy Tsyganok
Yaroslav Khrolenko
Iryna Domanetska
Olena Fedusenko
Julia Minaeva

The relevance of the work

- The influence of Internet technologies in almost all areas of human life and activity
- facilitating access to confidential information (private, commercial, state)
- Phishing is one of the most widespread types of cybercrime
- human factor
- improvement of phishing technologies

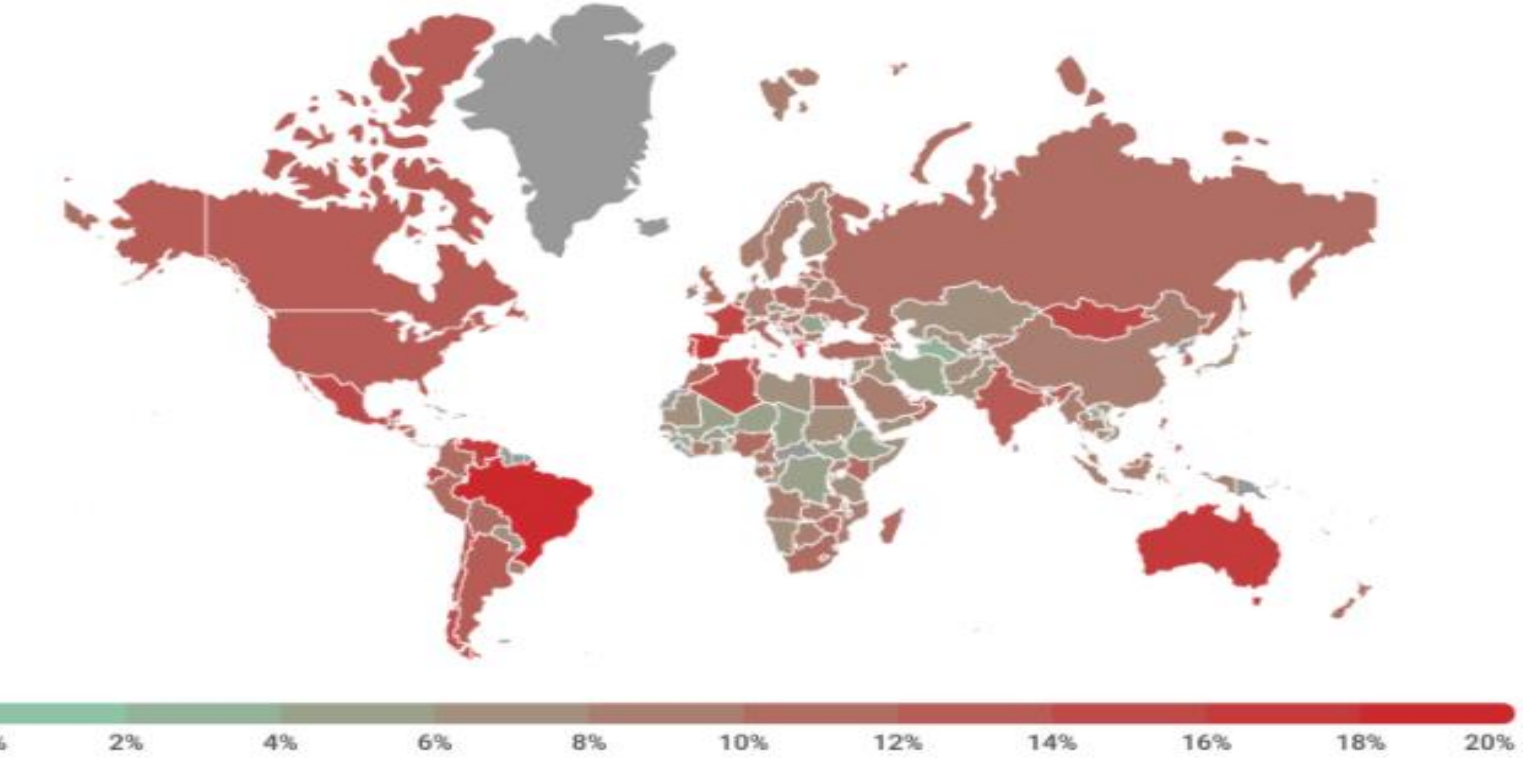
The purpose of the work

Increasing the effectiveness of protection against phishing attacks based on the use of neural networks

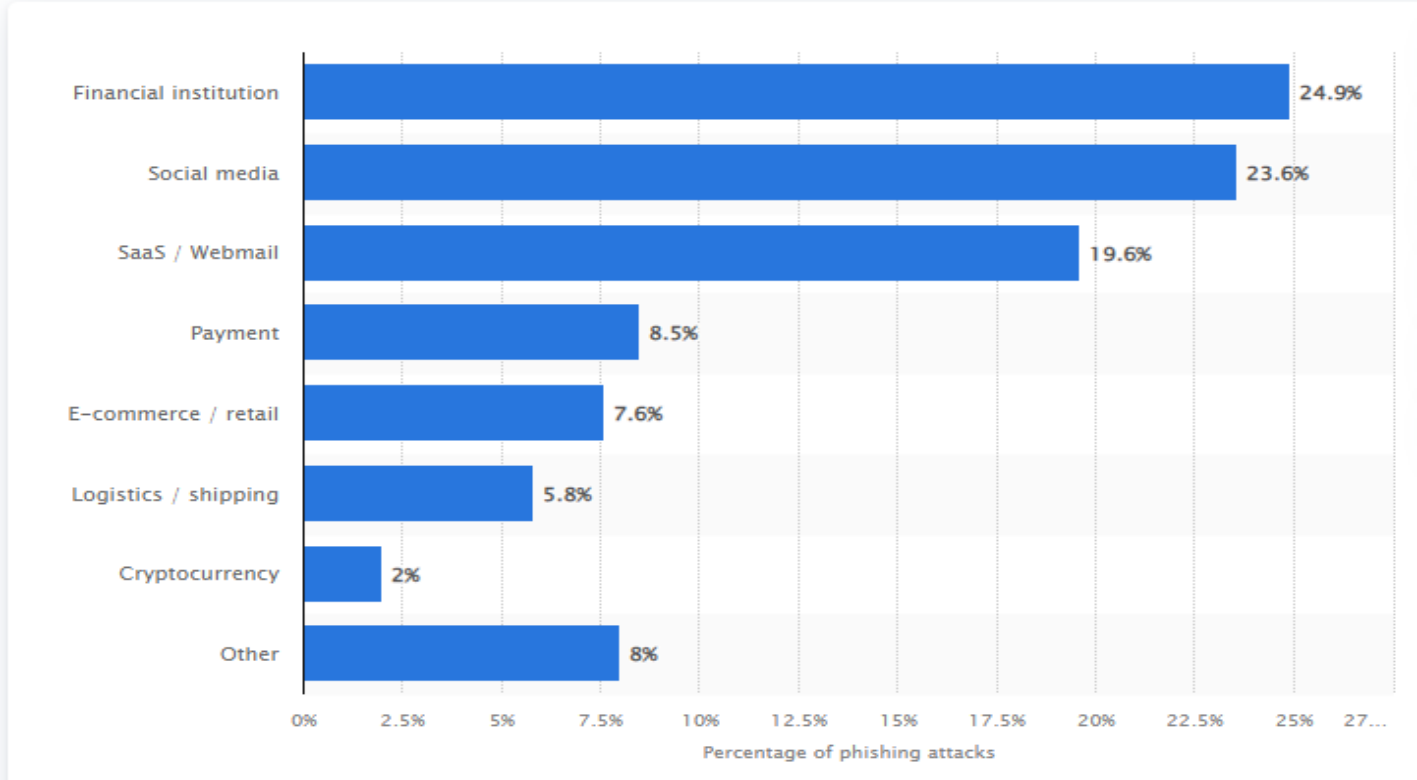
Object of research - The process of protecting against phishing attacks

Subject of research - Methods of protection against phishing attacks using neural networks

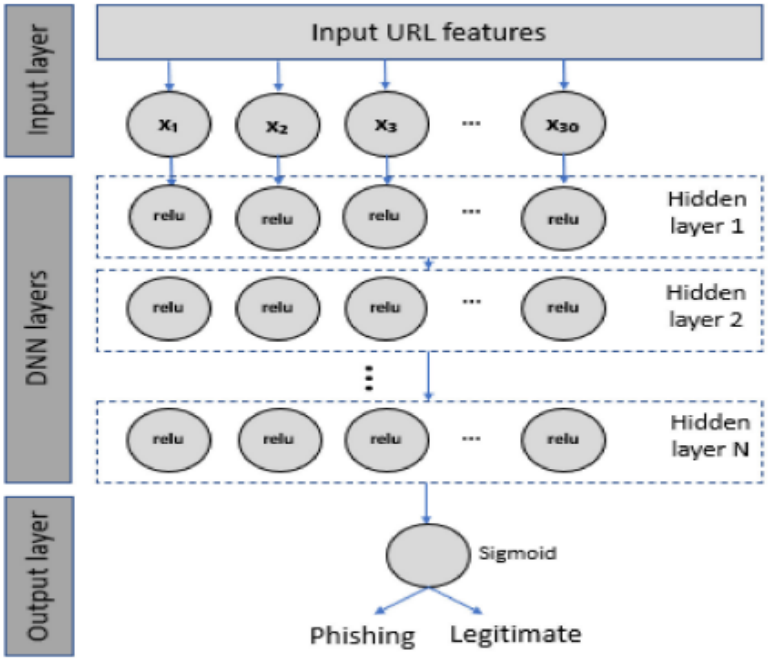
Phishing distribution and attack targets

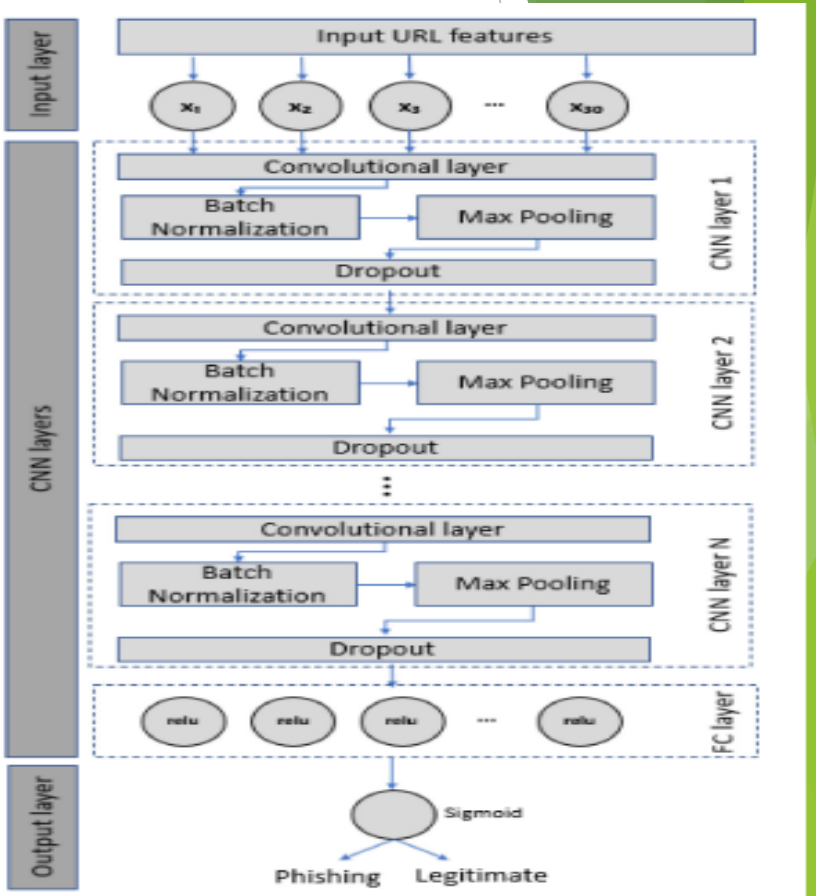


Phishing distribution and attack targets

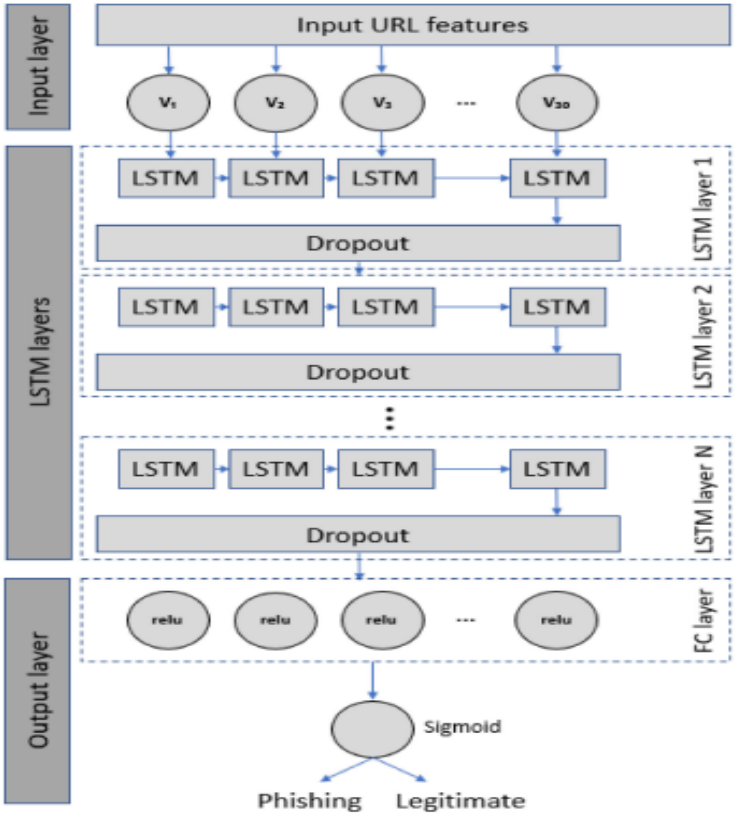


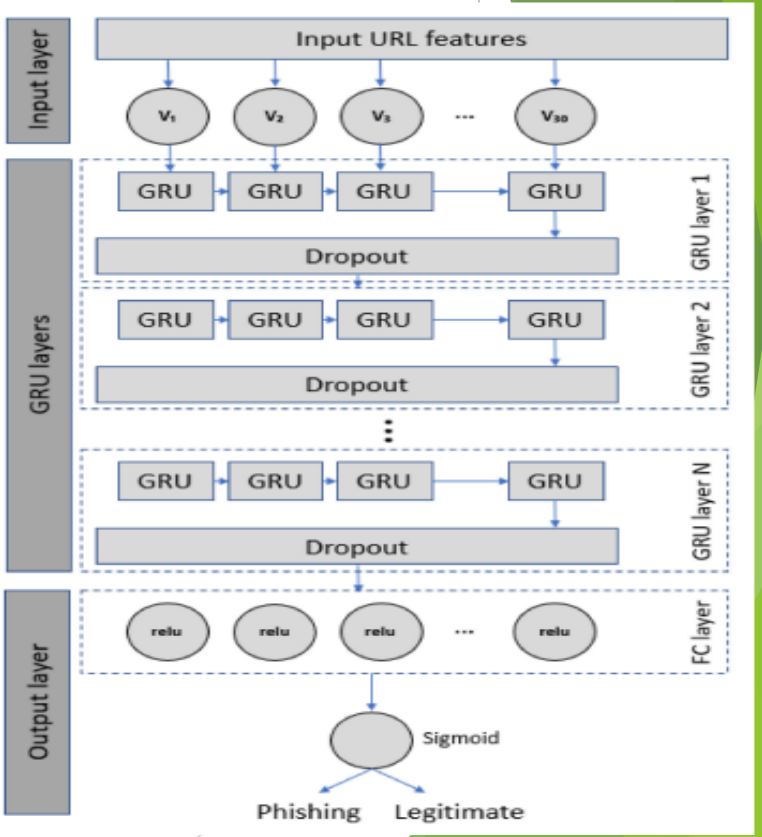
Analysis of neural network architectures used for phishing link recognition





Analysis of neural network architectures used for phishing link recognition





Dataset description

UCI Machine Learning Repository

10,000 copies

5000 phishing links

5000 legit websites

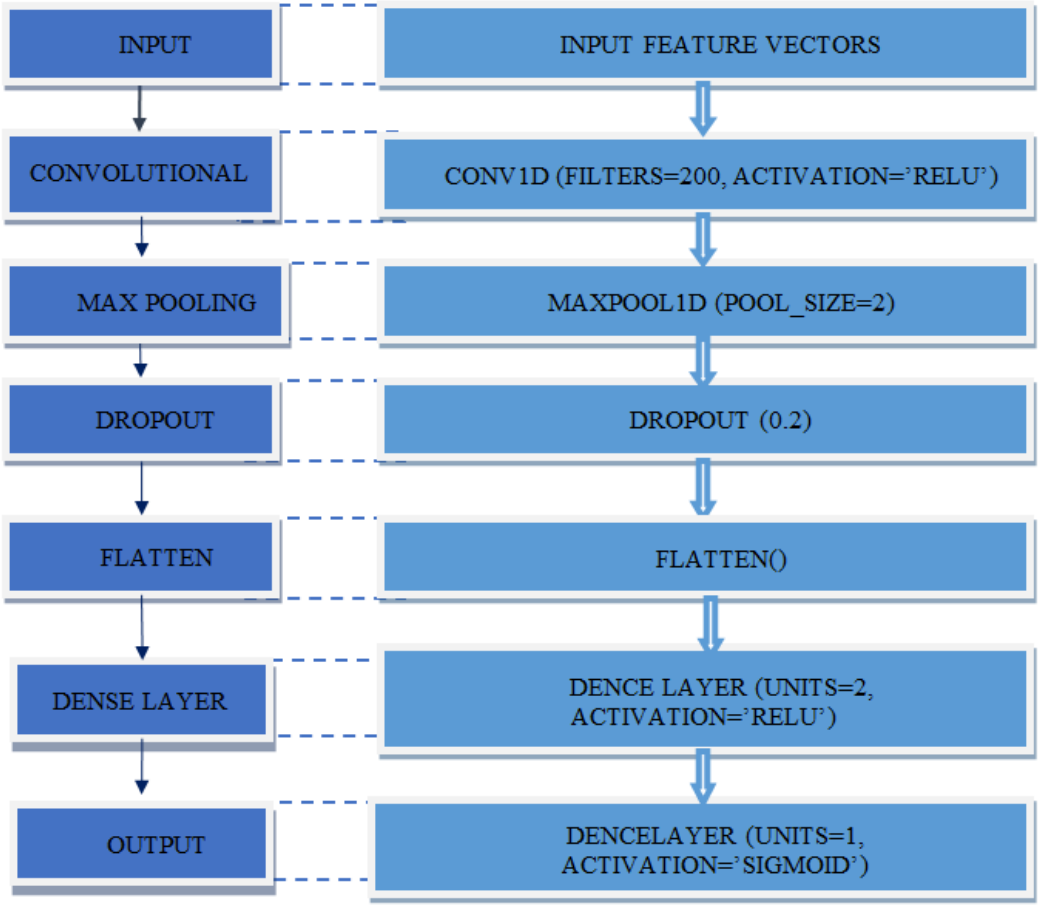
30 attributes

Attribute number	Attributes	Possible values
1	having_IP_Address	-1,1
2	URL_Length	1,0,-1
3	Shortening_Service	1,-1
4	having_At_Symbol	1,-1
5	double_slash_redirecting	-1,1
6	Prefix_Suffix	-1,1
7	having_Sub_Domain	-1,0,1
8	SSLfinal_State	-1,1,0
9	Domain_registration_length	-1,1
10	Favicon	1,-1
11	Port	1,-1
12	HTTPS_token	-1,1
13	Request_URL	-1,1
14	URL_of_Anchor	-1,0,1
15	Links_in_tags	1,-1,0
16	SFH (server form handler)	-1,1,0
17	Submitting_to_email	-1,1
18	Abnormal_URL	-1,1
19	Redirect page	0,1
20	onMouseOver (using to hide link)	1,-1
21	RightClick	1,-1
22	Using pop-up widnow	1,-1
23	Iframe	1,-1
24	age_of_domain	-1,1
25	DNSRecord	-1,1
26	web_traffic	-1,0,1
27	Page_Rank	-1,1
28	Google_Index	-1,1
29	Links_pointing_to_page	1,0,-1
30	Statistical_report	-1,1
Class	Result	-1,1

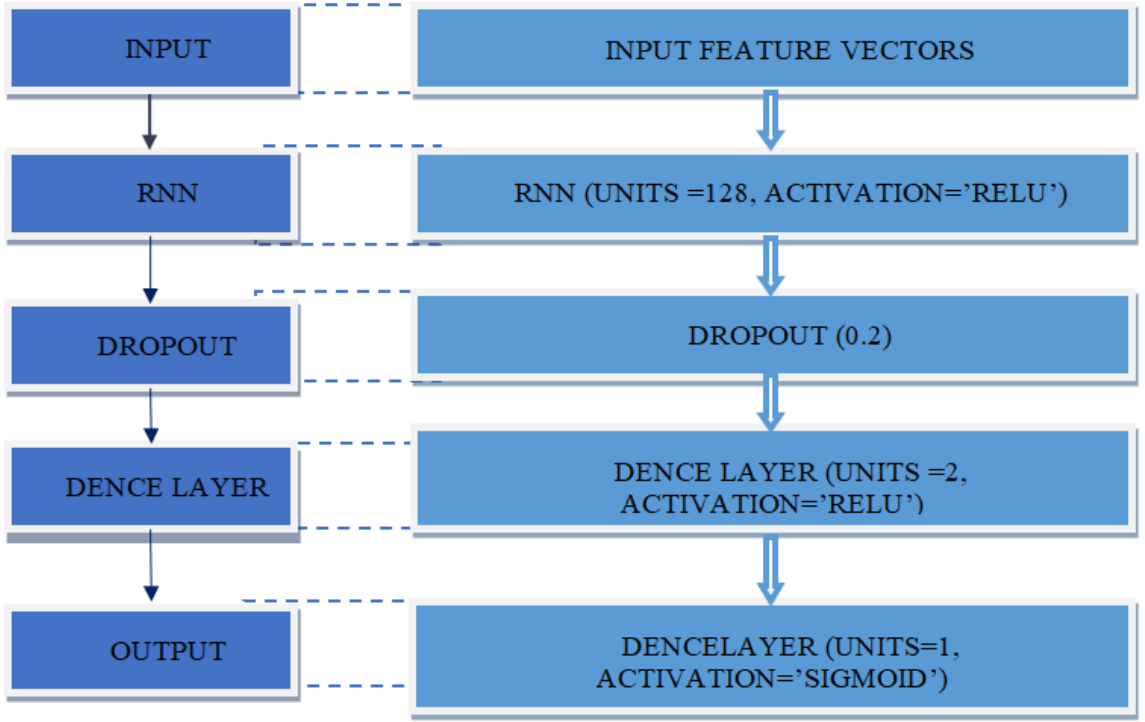
Comparing architectures

- Stage 1 - comparative analysis of networks built on a single basic model (CNN and RNN).
- Stage 2 - comparative analysis of the performance of the winning network of Stage 1 with the hybrid CNN-RNN network.
- Stage 3 - comparative analysis of the performance of the winning network of the 2nd stage with the hybrid DNN-LSTM network.

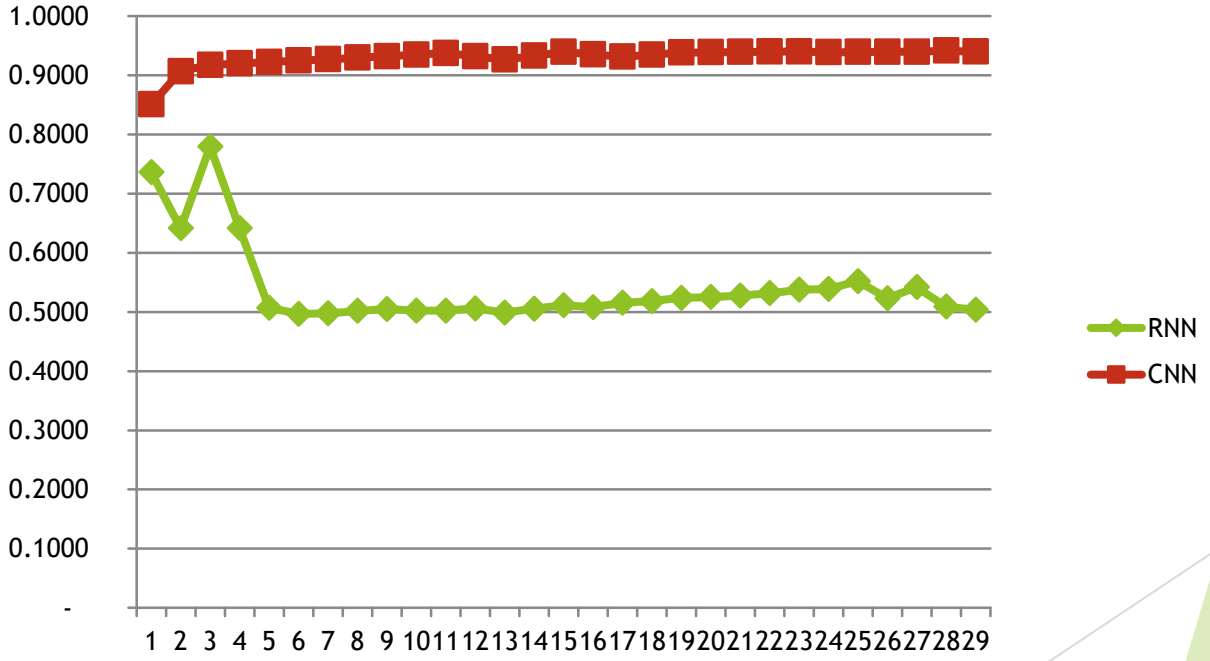
Architecture CNN



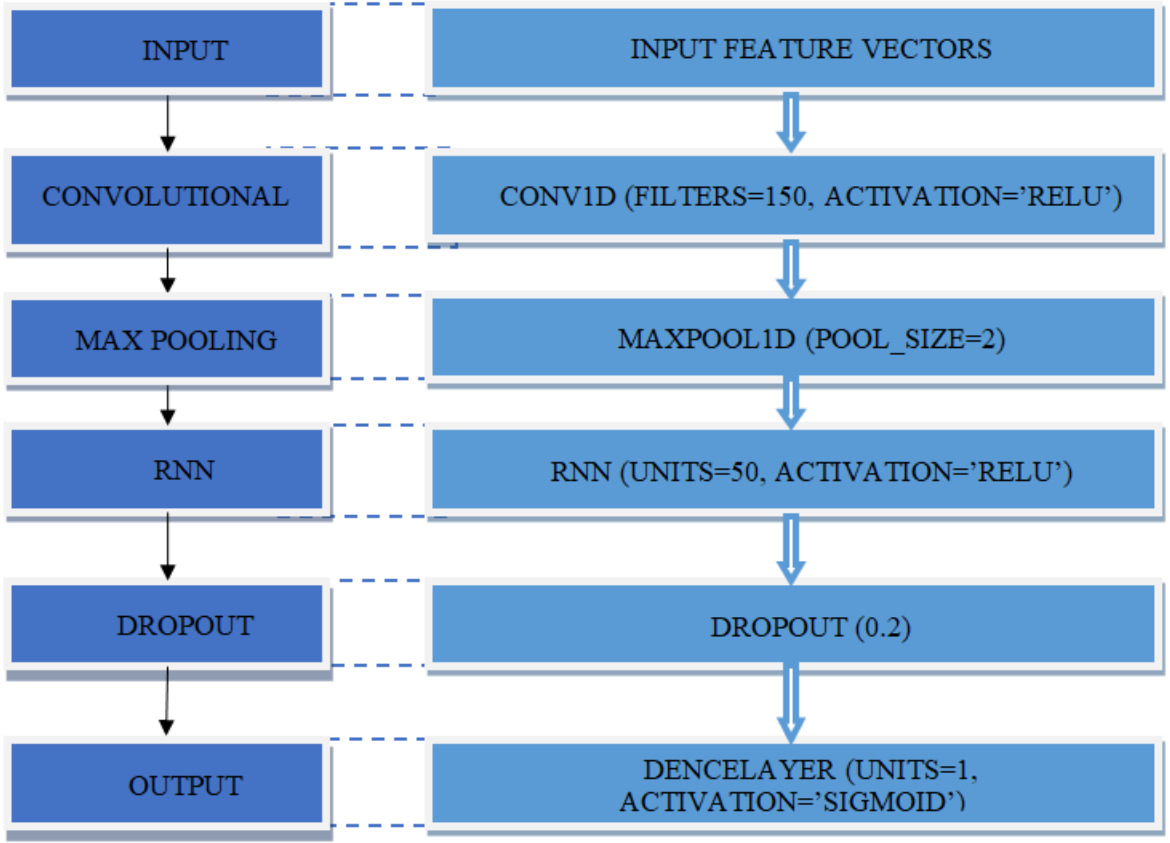
Architecture RNN



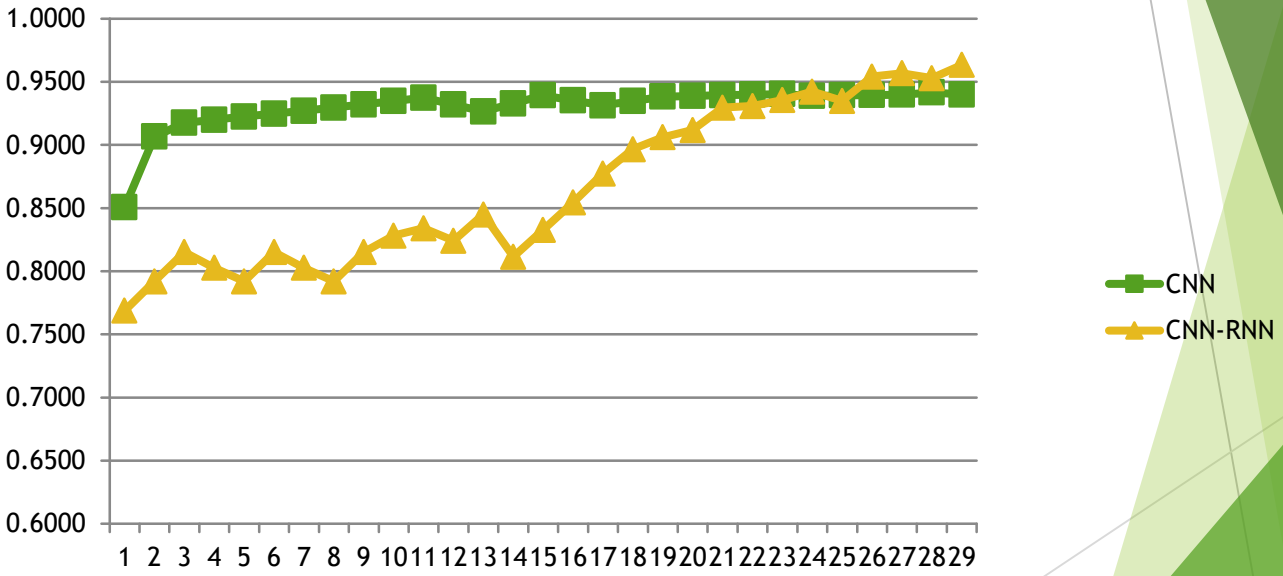
Architecture accuracy index for CNN and RNN systems



Architecture of hybrid convolutional recurrent network

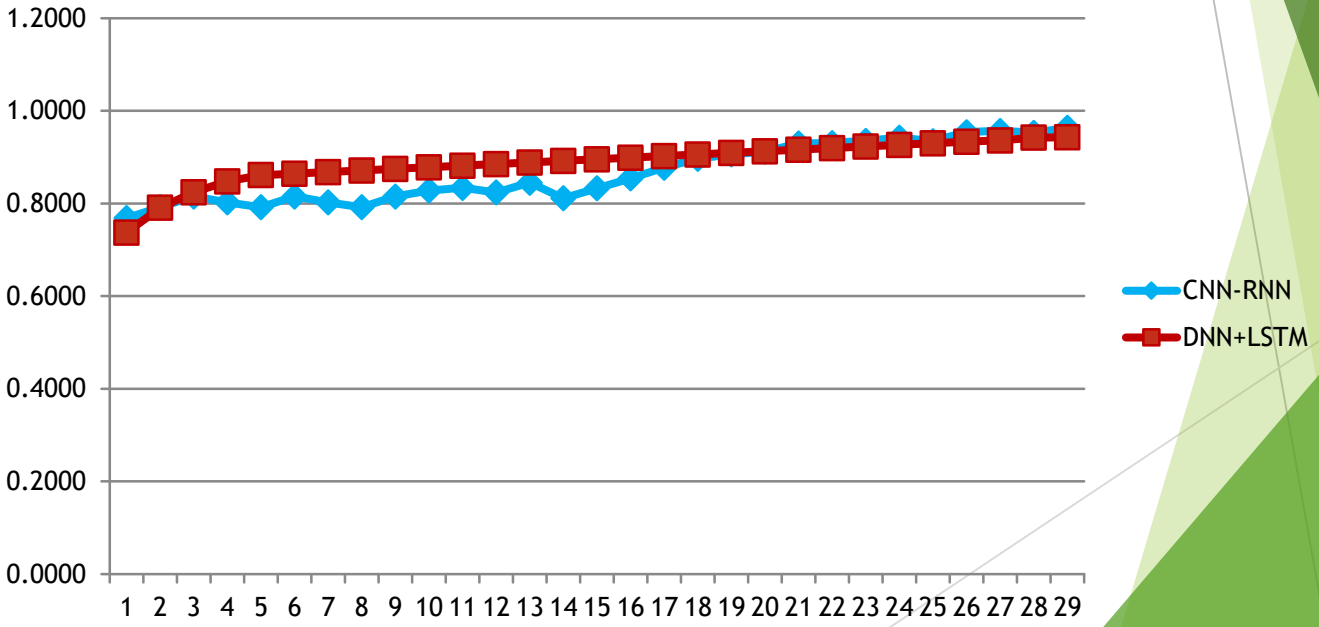


Accuracy indicator for CNN systems and the created hybrid network



Comparison with the DNN-LSTM system

№	Name	Performance parameters			
		loss	accuracy	precision	recall
1	**CNN-RNN**	0,1048	0,9634	0,9645	0,9584
2	**DNN-LSTM**	0,0921	0,9430	0,9452	0,9380



Summary

In the study

- an analysis of modern scientific publications devoted to the problem of anti-phishing based on neural networks was carried out
- perspective architectures of neural networks to solve the problem of increasing resistance to phishing attacks are defined
- a neural network of hybrid architecture to improve the effectiveness of anti-phishing tools was proposed

A promising area of future research is the development of neural network architectures using ensemble methods.