







# Predicting the Usability of the Dice CAPTCHA via Artificial Neural Network

Alessia Amelio<sup>1</sup> , Radmila Janković<sup>2</sup> , Dejan Tanikić<sup>3</sup> ,  
and Ivo Rumenov Draganov<sup>4</sup> 

<sup>1</sup> University of Calabria, DIMES, Rende, CS, Italy  
aamelio@dimes.unical.it

<sup>2</sup> Mathematical Institute of the S.A.S.A., Belgrade, Serbia  
rjankovic@mi.sanu.ac.rs

<sup>3</sup> University of Belgrade, Technical Faculty in Bor, Bor, Serbia  
dtanikic@tfbor.bg.ac.rs

<sup>4</sup> Technical University of Sofia, Department of Radio Communications and Video Technologies, Sofia, Bulgaria  
idraganov@tu-sofia.bg

**Abstract.** This paper introduces a new study of the CAPTCHA usability which analyses the predictability of the solution time, also called response time, to solve the Dice CAPTCHA. This is accomplished by proposing a new artificial neural network model for predicting the response time from known personal and demographic features of the users who solve the CAPTCHA: (i) age, (ii) device on which the CAPTCHA is solved, and (iii) Web use in years. The experiment involves a population of 197 Internet users, who is required to solve two types of Dice CAPTCHA on laptop or tablet computer. The data collected from the experiment is subject to the artificial neural network model which is trained and tested to predict the response time. The proposed analysis provides new results of usability of the Dice CAPTCHA and important suggestions for designing new CAPTCHAs which could be closer to an “ideal” CAPTCHA.

**Keywords:** Prediction · CAPTCHA · Usability

## 1 Introduction

A digital library in its broad meaning is a collection of digital items, which can include images, videos, documents, but also multimedia interfaces. In this sense, methods for exploring, processing, and analysing a collection of digital interfaces, e.g. the CAPTCHA interface, can be important for solving specific digital libraries issues.

CAPTCHA stands for Completely Automated Public Turing test to tell Computers and Humans Apart. This is a challenging test which is required to be solved by a human user in order to understand if he/she is a human or a computer robot (also called bot). In particular, if the human user is able to correctly

solve the test, then he/she is recognised as a human, otherwise he/she is considered as a bot [2].

The CAPTCHA test is currently used in multiple practical applications, including: (i) web systems for e-commerce, (ii) advanced authentication systems for e-mail, (iii) online pooling, (iv) different web systems, etc. [21].

The design of a CAPTCHA should follow some important requirements, such as: (i) the solution to the CAPTCHA should not depend on a given language which is known by the human user, (ii) the solution to the CAPTCHA should be independent from personal and demographic factors of the human user and should be given in no more than 30 s (postulate of “ideal” CAPTCHA [7]), (iii) the privacy of the human user should be preserved, (iv) the solution to the CAPTCHA should be easy for the human user and difficult for the bot [1].

In the last years, different types of CAPTCHA have been designed, which can be categorised as: (i) text, (ii) audio, (iii) image, (iv) video, and (v) interactive and puzzle CAPTCHA. The text and audio CAPTCHA require that a text or audio signal is recognised by the user and reported in a text field. They both proved to be insecure in different contexts since they could be solved by bots with Optical Character Recognition (OCR) and advanced speech processing algorithms. Accordingly, the CAPTCHA was equipped with image, video and interactive and puzzle tasks, for which the simulation of the human behaviour by a computer program is much more difficult. One of the last frontiers in CAPTCHA design is the interactive and puzzle area, which includes different CAPTCHA tests, e.g. FunCAPTCHA, SweetCAPTCHA as interactive ones and Dice CAPTCHA as a puzzle one [10]. In particular, in a puzzle CAPTCHA, the task is to solve a puzzle which can be equipped with images in order to discriminate a human user from a bot. Since the solution to the puzzle is quite difficult for a human subject, this CAPTCHA can take more time to be solved. Also, it is very difficult to be solved by a bot.

A CAPTCHA test can be analysed under different perspectives, which include: (i) security, (ii) practicality, and (iii) usability [4]. The security is mainly connected to the robustness of the CAPTCHA test to attacks which are made by bots. This represents the main concern of the CAPTCHA designers. The practicality includes the main aspects related to the CAPTCHA programming. Finally, the usability analyses the CAPTCHA test under a user-centric perspective. This represents the main aspect related to the use of the CAPTCHA, which includes the interaction of the user with the test.

## 2 Related Work

In the last decades, different works have been proposed in the literature concerning the usability of the CAPTCHA test.

Yan and El Ahmad [22] designed a framework for the exploration of the CAPTCHA usability. They relate their analysis to the following components: learnability, memorability, efficiency, errors and satisfaction. The latter three are

determined by the following usability criteria: accuracy, response time and perceived difficulty, further expanded by the authors to distortion, content and presentation in a 3-dimensional framework. Evaluating various types of CAPTCHAs with it, led to the following general conclusions: (i) foreigners may be hampered in solving the text-based CAPTCHA, (ii) the predictability level of text sequences has the major effect on usability and security, as well as the use of the colour.

To further enhance the security of the text-based CAPTCHAs, Lupkowski and Urbański [16] proposed the users not only to recognise but also to understand the presented text. Introducing semantic priming, the researchers have proven that there is a significant dependence of the response time based on the users' experience with the newly proposed system. In the same time, the decrease of the response time does not influence the accuracy of solving the CAPTCHA. A subjective evaluation of the difficulty in solving it from the user's point of view was carried out on a 10-level scale. Less experienced users reported slightly higher levels of difficulty and that result seems to be highly correlated with the registered higher response times.

In [13], Ince et al. estimated the execution times of the interactive 3D CAPTCHA using a keystroke level model. Various system usage scenarios were tested. This proved possible to predict the necessary time of accomplishing specific tasks by the users. It includes up to 8 distinctive steps to perform all registered by their execution times. Upper limits for them are found along with the average action time for a complete solving of the CAPTCHA. Three types of errors are detected during the experimentation with their estimated frequency occurrence. However, no statistical analysis is performed against the users' response times.

Users' experience and solution performance on the reCAPTCHA system are statistically investigated more deeply by Beheshti and Liatsis [5]. Taking into account a 3 level model including distortion, content and presentation, the authors conducted a user survey with 13 questions. They considered gender, sex, age, glasses worn by the users, and the type of used monitor. The quantitative performance is represented by the number of needed attempts for a successful solution, the time needed for it along with the used character size and lengths. Some subjective measures announced by each user are the level of willingness to solve similar tasks again, the difficulty in recognising symbols and the experienced ambiguity level. Also, the personal preference on the language of the text and the level of interaction with the system were measured. No precise relation was declared among any of these parameters of the study and no attempt was made to predict the performance of the users in any future activity based on their properties.

Alsuhibany [3] tested a recently introduced optimiser within the CAPTCHA employing the collapsing mechanism over the text to find the effect on the users' performance. An improvement is thought to be statistically significant for both the solution accuracy and time. Age, sex and affiliation background are taken into account during the experiments. The experimental results include average accuracy and solution time with common statistical derivatives before and after

the optimisation. Complete distributions among all users of these parameters are also presented. A subjective level of difficulty in solving the CAPTCHA prior and after the improvement is also provided as a feedback by the users. A considerably higher personal comfort is reported after the optimisation. No relation was presented between the users' properties and the achieved performance.

A significantly more complete study [18] over the usability of the CAPTCHA, the Dynamic Cognitive Game (DCG) one, in particular, and its relation to stream relay attacks is made by Nguyen. Presenting a customised version of that system, the author undertook a test with 40 participants, primarily students, knowing their demographics. A scale from 1 to 5 was used to evaluate the subjective user experience. Additional questions were asked for a users' feedback. Completion time, non-successful completion part of all users, effects of objects number and speed in the test on the response time and accuracy were the objective parameters to register. Positive results in performance are reported from using the proposed system's enhancement but no relation with the users' properties has been investigated.

Extensive results on testing the DCG CAPTCHA are given in [17] by Mohamed et al. Exploring the usability and security of the system in terms of completely automated attacks, relay attacks based on human interaction and hybrid types of interference, reveal that it remains usable and stays resilient in the first case, no matter of the access device type, stationary or mobile. Some vulnerability is discovered for the latter two groups of attacks. Demographics of all 40 participants in the test were taken into consideration, in particular the gender, age and education. Response time, success rate and subjective user experience were measured and statistically processed but no correlation to the users' demographics has been shown.

Another investigation on the usability of the CAPTCHA is presented in [23] concerning the application of Chinese characters in the tasks presented to the users. An analytical comparison is performed over test-cases including common alphanumeric labels and Chinese ones apart trying to get a deeper understanding of the cognitive processes involved during the solution. Only the age of the participants in the tests has been considered. Software and hardware properties, such as display size, of the testing platforms were also taken into account. Extensive statistical parameters from the processing of the test data are presented which prove the comparable level of usability of both CAPTCHAs and useful guidelines of designing a Chinese system were derived. No relation to the users' age or other demographic features was presented. By contrast, the age discrimination on the user performance was studied by Guerrar et al. [11] for CAPTCHAs solved on mobile devices. The dependency of the response time, accuracy, pleasantness, solvability, input mechanism type, understandability, suitability, memorability, and preference from distinctive age ranges is presented from which useful guidelines could be given for the future use of CAPTCHAs based on that user feature. All statistical analysis of the data presented in the aforementioned works could be unified by the introduction of a recently proposed robust metric for the comparison based on a multiagent system modelling [12].

In different works [6–8], Brodić et al. presented results from a recent study which aims to investigate the CAPTCHA usability not only from a point of view of the supporting software and hardware adopted by the users but also based on a variety of demographic features which describe them. In one of the tests, a user-centric approach is presented for which 190 participants were gathered to solve the Dice CAPTCHA on both laptops and tablets [7]. Age, gender and education level were taken into consideration during a statistical derivation of the related dependence of the response time. General conclusions were drawn about the usability of the Dice CAPTCHA and its closeness to an “ideal” CAPTCHA. An attempt of predicting the response time to solve image and interactive CAPTCHAs was made in [6] based on age, education level and Web use level of a population of 114 users. A regression tree was used to evaluate the prediction accuracy which appeared to be high enough for practical purposes of future CAPTCHA designs targeted to specific user groups. An advanced statistical analysis was also implemented based on a study with 197 subjects registered with their age and Web use level solving the Dice CAPTCHA on a tablet or laptop [8]. It was based on association rule mining with the response time and the number of solution tries as dependent variables. The co-occurrence of factors affecting the Dice CAPTCHA usability may be established using this approach which helps in understanding which type of Dice CAPTCHA is closer to an “ideal” CAPTCHA.

In this paper, we propose an artificial neural network model which could be trained and then tested to predict the Dice CAPTCHA response time from known characteristics of the users who solve the CAPTCHA, i.e. age, Web use in years and used device type. In comparison to the previous works [6–8], we propose the following novelties: (i) a predictability analysis on a different type of CAPTCHA tested with a different prediction model, and (ii) an extension of the usability analysis of the Dice CAPTCHA via predictability of its response time given known users’ characteristics. It will provide new results on the usability of the Dice CAPTCHA and useful insights for designing future CAPTCHAs which could be closer to the “ideal” CAPTCHA.

The rest of the paper is organised as follows. In Sect. 3 the Dice CAPTCHA is described with its two used types. Then, in Sect. 4, the developed artificial neural network model is presented, followed by a description of the experimental setup in Sect. 5. Experimental results are given and discussed in Sect. 6. At the end, in Sect. 7 conclusions are drawn.

### 3 The Dice CAPTCHA

Dice CAPTCHA is a puzzle-based CAPTCHA where the user needs to roll a dice and enter the numbers which are visualised on the faces of the dice, or their sum in a given field [10]. If the user correctly solves the Dice CAPTCHA, then he/she will be considered as a human, otherwise he/she will be considered as a bot and his/her request to access a web form will be denied.

There are two types of Dice CAPTCHA: (1) Homo-sapiens Dice (Dice 1), where the challenge of entering the sum of the digits shown on the faces of the

dice is presented to the user (see Fig. 1 (a)), and (2) All-the-rest Dice (Dice 2), where the user needs to enter the exact numbers as they are presented on the dice in a given field (see Fig. 1 (b)).



**Fig. 1.** (a) Dice 1 CAPTCHA (Homo-sapiens Dice), (b) Dice 2 CAPTCHA (All-the-rest Dice)

Dice CAPTCHA is designed with a variety of colour skins as presented in Fig. 2. In this case, it is easy to presume that the colour can have an effect on the users' response time of solving the CAPTCHA. When designing colour CAPTCHAs, it is assumed that different colour schemes can serve as a defence against attacks, in particular attacks made by an OCR software [22]. However, the use of the colour when designing CAPTCHAs should be taken with caution, as adding colour can sometimes have a negative impact on usability and security [22]. Interested parties that would like to use the Dice CAPTCHAs for protecting their websites from the bots can choose the number of the dice their Dice CAPTCHA will consist of.



**Fig. 2.** Dice CAPTCHA colour skins

## 4 The Artificial Neural Network Model

An Artificial Neural Network (ANN) model is used for predicting the response time to solve the Dice 1 and Dice 2 CAPTCHAs from known personal and

demographic characteristics of the users who solve the CAPTCHA. The response time can be influenced by many factors. From the previous study [8], the age of the user, the device type on which the CAPTCHA is solved and the Web use (in years) of the user are marked as the main influencing factors. Consequently, they will be used in this study.

ANN is a structure consisting of a large number of neurons, which are organised in a few number of layers. The layers of neurons are fully interconnected so that each neuron in the current layer of neurons is connected with all neurons from the previous one, as well as all neurons in the following layer. ANN has an input layer with one neuron for each input value and an output layer with one neuron for each output value. In this case, there will be 3 neurons in the input layer of the ANN, which represent: (i) age, (ii) device type and (iii) Web use. Also, ANN has just one output neuron in the output layer which represents the CAPTCHA response time (see Fig. 3 (a)).

The number of hidden layers and the number of neurons in each of them is not known in advance. But the processing ability of the ANN depends on the number of layers of hidden neurons and their number. Although there exist some heuristic methods for determining this number, the only certain method for resolving this problem is the trial-and-error approach. It is a critical design parameter because the performance of the ANN may not be adequate with an insufficient number of layers and neurons, while a large number of layers and neurons causes poor generalisation on the new data [20].

This kind of ANNs is sometimes called Feedforward Artificial Neural Network, because the information only travels forward (with no loops), through the input nodes, then through the hidden nodes and finally through the output nodes.

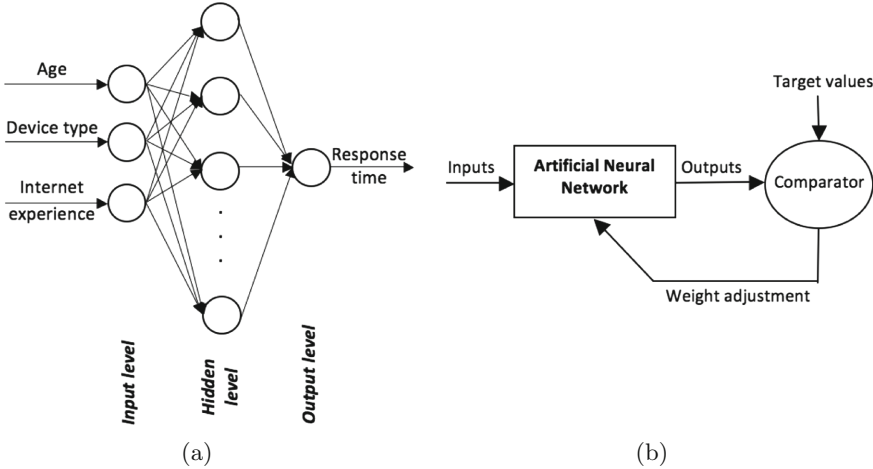
The performance of one neuron is pretty simple. It sums up the signals from the neurons which lay in the previous layer multiplied with interconnection weights and biases, forming in that way the neuron potential. The activation function produces the output from the neuron, according to the potential of the neuron. The activation function, also known as transfer function, invokes non-linearity into the ANN, which enables in that way modelling very complicated relationships among input and output parameters. This can be represented by the following equations:

$$a_i = \sum_{j=1}^n w_{ij} \cdot x_j + b_i, \quad y_i = f(a_i) \quad (1)$$

where  $a_i$  is the potential of the  $i$ -th neuron,  $w_{ij}$  is the adjustable interconnection weight between the  $j$ -th neuron in the previous layer and the  $i$ -th neuron in the current layer,  $x_j$  is the output from the  $j$ -th neuron serving as input into the  $i$ -th neuron,  $n$  is total number of inputs into the  $i$ -th neuron,  $y_i$  is the output from the  $i$ -th neuron, and  $f$  represents the activation function (transfer function).

The ANN has to be trained according to a training data set, i.e. known input/output pairs. The training data set must be representative, so that the desired accuracy of the ANN could be achieved. The training process is initialised

by assigning random values to weights and biases. The input values are presented to the ANN, so that it can calculate the output value. The difference between the desired and calculated value represents the error, which has to be reduced. The values of the weights and biases are changed in that way to minimise the overall error. The neural networks are usually trained to provide targeted outputs for specific inputs. The network is adjusted, based on the comparison between outputs and target values, until the match between the outputs and target values is satisfactory. The algorithm of the ANN training is shown in Fig. 3 (b).



**Fig. 3.** (a) Architecture of the used ANN, (b) Algorithm of ANN training

## 5 Experiment

The aim is to determine if the response time of solving the Dice CAPTCHA can be efficiently predicted based on users' personal and demographic features. In that case, the investigated CAPTCHA cannot be considered as an “ideal” one, since its response time can be easily predicted from users' features.

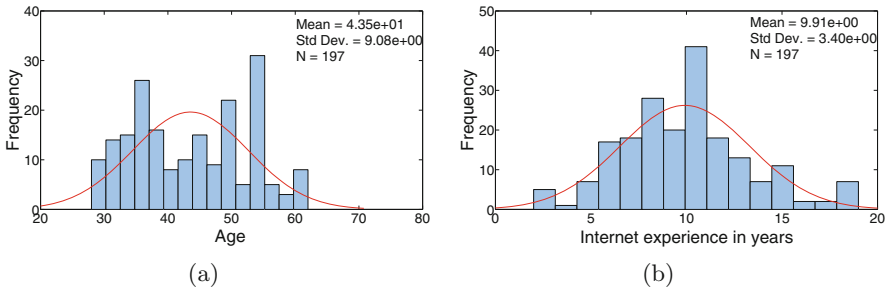
The experiment involves a population of 197 Internet users with the Dice 1 and Dice 2 CAPTCHAs to solve on a laptop and tablet computer. The laptop used for the experiment has the following characteristics: (i) 15.6” wide screen, (ii) CPU Quad-core 2.4 GHz, (iii) 4 GB of RAM, (iv) 500 GB of internal memory, and (v) operating system Microsoft Windows 7. By contrast, the tablet used for the experiment has the following characteristics: (i) 7” wide screen, (ii) CPU Quad-core 1.2 GHz, (iii) 1 GB of RAM, (iv) 16 GB of internal memory, and (v) operating system Android.



## 5.1 Data Gathering and Dataset Creation

The users voluntarily participated in this study and agreed that their data would be anonymously used for research purposes. They were involved through personal email. Before starting the experiment, the users gave their consent through an online form. In order to avoid bias effects, the participants were not informed about the aim of the study. For each user, the response time to correctly solve both Dice CAPTCHAs was measured in seconds, from the beginning of the task until the end of the task. The measured times were then registered in a dataset. Accordingly, there are 197 instances consisting of 4 variables: (1) age, (2) type of device where the CAPTCHAs are solved (tablet or laptop), (3) Web use in years, and (4) response time.<sup>1</sup>

From these 197 users, 100 users solved the CAPTCHA on a tablet computer, while 97 users used a laptop computer for solving the CAPTCHA. Around 62% of the users were male, while the rest of 38% were female. The age of the users who were involved in the experiment ranges between 28 and 62 years, while the Web use ranges between 1 and 19 years. Figure 4 shows the distribution of the values of the ages and Web use of the participants.



**Fig. 4.** Distribution of the values of (a) ages and (b) Web use of the participants [8]

From Fig. 4 (b), it is worth noting that the distribution of the Web use has a Gaussian like shape [8]. By contrast, the distribution of the ages in Fig. 4(a) is deviant from the ideal normal distribution [8], with the highest frequency between 32 and 40 years, and for 50 and 54 years.

Considering the response time, the users solved the Dice 2 faster than the Dice 1 CAPTCHA. Also, the mean value of the response time for the Dice 1 CAPTCHA is 9.48 s, and for the Dice 2 CAPTCHA is 7.34 s. The median values are 8.00 s and 6.00 s for the Dice 1 and Dice 2 CAPTCHAs, respectively.

<sup>1</sup> The gathered data is freely available at: <https://sites.google.com/site/alessiaamelio/software-tools/dice-captcha-dataset>.

## 5.2 Experimental Setting

Before the analysis, the collected values for the measured variables have been normalised in the range  $[0,1]$ , using the following equation:

$$x_1 = \frac{x - \min_x}{\max_x - \min_x} \quad (2)$$

where  $x_1$  is the normalised value,  $x$  is the actual value, and  $\min_x$  and  $\max_x$  are the minimum and maximum values for the variable associated to  $x$ , respectively. After the normalisation, the prediction of the response time was made using the ANN. Then, the predicted values were denormalised using the inverse formula of Eq. (2) and compared to the actual response time values.

The number of layers and type of activation function were varied in the ANN. Since no meaningful changes were obtained, the simple ANN architecture  $3-N_h-1$  was selected, where  $N_h$  represents the number of the neurons in the hidden layer. Also, a trial-and-error approach was carried out for adjusting the number of neurons in the hidden layer. In particular,  $N_h$  was varied from 5 to 50 with intervals of 5 (5, 10, 15, 20, etc.). In that way, 10 different ANNs have been created and tested.

In order to check the ANN performance, some data was used for training, some for testing and the rest of data for ANN validation. The sizes of the adopted data sets are: 75% of instances for training, 10% of instances for validation and 15% of instances for testing. It is worth to say that the training, validation and test sets were varied, too. The adopted algorithm is the Levenberg–Marquardt algorithm [15], and the maximum number of epochs is 1000. The training of the ANN is repeated until the error function (in this case Mean Squared Error – *MSE*) between the predicted values and the desired (target) values is minimised.

In order to evaluate the prediction accuracy, the following measures are used: (i) Pearson’s correlation coefficient [14] –  $R$ , which is computed between the desired (target) and predicted values, (ii) Error, which is the difference between the desired (target) and predicted values, and (iii) *MSE*.

## 6 Results and Discussion

The experimentation has been performed in Matlab R2017a and Weka version 3.7 on a laptop with Quad-core CPU 2.2 GHz, 16 GB of RAM and Unix operating system.

In the following, the results in terms of  $R$  coefficient of the trial-and-error procedure for detecting the best number of neurons in the hidden layer are discussed for Dice 1 and 2 CAPTCHA. Then, the prediction accuracy of the tuned ANN is analysed in terms of Error and  $R$ . Since close results are obtained by different combinations of training, validation and test sets, they will be reported for only one of these combinations. Finally, comparison results with other regression methods in terms of *MSE* and  $R$  are shown and discussed.

### 6.1 Analysis of the Hidden Layer

Table 1 reports the results of the trial-and-error procedure in terms of  $R$  coefficient for Dice 1 and 2 CAPTCHA on the test set. It is worth noting that the highest values of  $R$  correspond to hidden layer composed of 25 neurons for Dice 1 CAPTCHA, and 45 neurons for Dice 2 CAPTCHA. This same number of neurons corresponds to low  $MSE$  values of 0.02 and 0.04 for Dice 1 and 2 CAPTCHA, respectively.

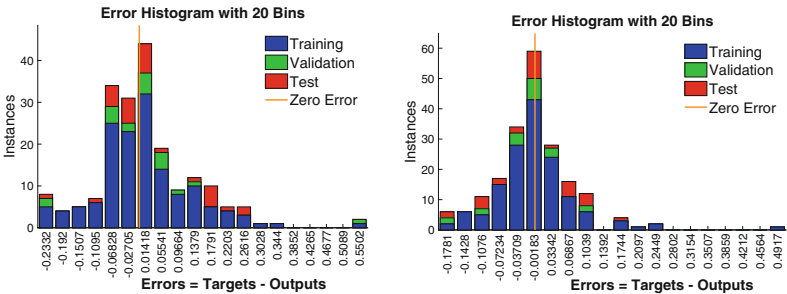
**Table 1.** Pearson’s correlation coefficient  $R$  on test set for Dice 1 and 2 CAPTCHA. The number of neurons in the hidden layer is varied from 5 to 50. The best values are marked in bold

	5n	10n	15n	20n	25n	30n	35n	40n	45n	50n
Dice 1	0.711	0.523	0.667	0.446	<b>0.804</b>	0.431	0.458	0.403	0.653	0.783
Dice 2	0.246	0.270	0.485	0.343	0.227	0.076	0.369	0.033	<b>0.542</b>	0.204

Accordingly, the ANN model with hidden layer composed of 25 neurons for Dice 1 CAPTCHA and 45 neurons for Dice 2 CAPTCHA will be in the focus for further analysis.

### 6.2 Analysis of the Prediction Accuracy

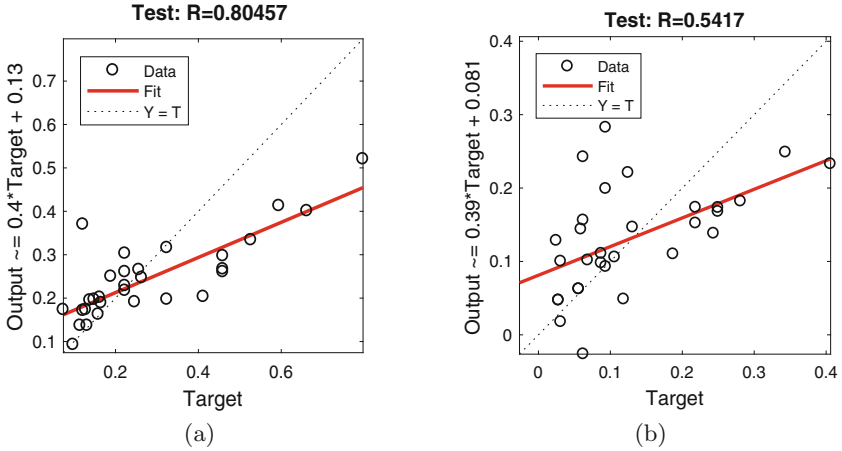
Figure 5 shows the histogram of the Error for Dice 1 and Dice 2 CAPTCHA. It is worth noting that both errors are far from zero. In the test set of Dice 1, most of instances show an error between  $-0.07$  and  $0.01$ . By contrast, in the test set of Dice 2, more instances are distributed in a larger error range between  $-0.11$  and  $0.10$ .



**Fig. 5.** Histogram of the Error for Dice 1 (left), and Dice 2 CAPTCHA (right)

In Fig. 6, the regression plots illustrate the network outputs with respect to target values in the test set for Dice 1 and 2 CAPTCHA, respectively. In order

to obtain a perfect fit, the data should fall along a  $45^\circ$  line, which indicates that the network outputs are equal to the target values. It is worth noting that the fit is worst in Dice 2, with lower values of  $R$  than Dice 1 CAPTCHA (0.54 for Dice 2 vs. 0.80 for Dice 1).



**Fig. 6.** Regression plots for (a) Dice 1 CAPTCHA (25 neurons in the hidden layer), and (b) Dice 2 CAPTCHA (45 neurons in the hidden layer)

From the results obtained by ANN, it is worth noting that the response time of both Dice 1 and 2 CAPTCHA is not perfectly predictable from known users' personal and demographic features: (i) age, (ii) type of device on which the CAPTCHA is solved, and (iii) years of Web use. In particular, the response time of Dice 2 is less predictable than Dice 1 CAPTCHA (higher Error and  $MSE$ , and lower  $R$ ). Also, it is worth saying that Dice 2 CAPTCHA has an average and median response time which is less than 30 s. Still, from the previous study [8], Dice 2 was less dependent on the users' features than Dice 1 CAPTCHA. These results confirm that Dice 2 tends to be closer to an "ideal" CAPTCHA than Dice 1.

### 6.3 Comparison Results

The results obtained by ANN in terms of  $MSE$  and  $R$  coefficient are compared with results obtained by other two well-known regression methods: (i) Regression Trees (RT) [19], and (ii) Support Vector Machine Regression (SVMR) [9].

In Weka, SMOReg and REPTree algorithms with a 10-fold cross validation were applied on the dataset in order to predict the response time of Dice 1 and Dice 2 CAPTCHAs based on age, type of device, and Web use in years.

The SVMR is implemented in Weka through an improved SMOReg algorithm which finds the best fitted line that minimizes the cost function error.

The instances from the training set that are closest to that line are called support vectors. The training data was normalized during the application of the algorithm. A complexity parameter is set to 1 which means that the violations of the margin are allowed, and the polynomial kernel with exponent 1 was used, which makes it a linear kernel. Also, REPTree (Reduced Error Pruning Tree) uses different iterations to create multiple regression trees, after which the algorithm selects the best generated tree. The splitting criterion is the information gain, while reduced error pruning is used as a criterion for pruning the tree. The maximum tree depth was set to no restriction. The algorithm also used 3 folds for pruning the tree of Dice 1 CAPTCHA, and 4 folds for Dice 2 CAPTCHA.

Table 2 shows a comparison of the prediction results.

**Table 2.** Comparison results of response time prediction for Dice 1 and 2 CAPTCHA

	REPTree		SMOreg		ANN	
	<i>MSE</i>	<i>R</i>	<i>MSE</i>	<i>R</i>	<i>MSE</i>	<i>R</i>
Dice 1	20.25	0.51	20.69	0.51	0.02	0.80
Dice 2	12.37	0.27	11.93	0.31	0.04	0.54

It is worth noting that the results obtained by RT (REPTree) and SVMR (SMOreg) are consistent with ANN, since the response time is not perfectly predicted by the adopted users' features. In particular, it is visible that the response time of Dice 2 is less predictable than Dice 1 in terms of *R* coefficient (0.27 vs. 0.51 for REPTree and 0.31 vs. 0.51 for SMOreg). However, ANN proved to be the most reliable method for predictability analysis, since it obtains the best performances in terms of *R* and *MSE*.

## 7 Conclusions

This paper extended the study of the Dice CAPTCHA usability by analysing the predictability of its response time given known users' personal and demographic characteristics. This was accomplished by proposing a new ANN model for the evaluation of the prediction accuracy. According to the postulate of "ideal" CAPTCHA (response time should not depend on personal or demographic factors of solving users), a low predictability of time to complete implies a better quality of a CAPTCHA. The main result of this study is that response time of both Dice CAPTCHAs is not perfectly predictable from users' features. In particular, the response time to solve the Dice 2 is less predictable than Dice 1 CAPTCHA from the users' features. This implies that Dice 2 is closer to an "ideal" CAPTCHA than Dice 1.

Considering the difference between Dice 1 and Dice 2 and the features of the Dice CAPTCHA, some useful suggestions can be made for designing new CAPTCHAs which can be closer to an “ideal” one. They are the following: (i) visualisation is to prefer to calculation of some result in the CAPTCHA design, (ii) it is preferred to split the task in smaller steps (like in the Dice 2 where each digit must be recognised and reported).

Since there is not much literature about the effects of the colour of the CAPTCHA on its response time, future work will include testing the usability of different coloured Dice CAPTCHAs.

**Acknowledgments.** This work was partially supported by the Mathematical Institute of the Serbian Academy of Sciences and Arts (Project III44006). The authors are fully grateful to the participants to the experiment for anonymously providing their data. This paper is dedicated to our colleague and friend Associate Professor Darko Brodić with full gratitude.

## References

1. von Ahn, L., Blum, M., Hopper, N.J., Langford, J.: Captcha: using hard AI problems for security. In: Biham, E. (ed.) *Advances in Cryptology - EUROCRYPT 2003*. LNCS, vol. 2656, pp. 294–311. Springer, Berlin Heidelberg (2003). [https://doi.org/10.1007/3-540-39200-9\\_18](https://doi.org/10.1007/3-540-39200-9_18)
2. von Ahn, L., Blum, M., Langford, J.: Telling humans and computers apart automatically. *Commun. ACM* **47**(2), 56–60 (2004). <https://doi.org/10.1145/966389.966390>
3. Alsuhibany, S.A.: Evaluating the usability of optimizing text-based captcha generation methods. *Int. J. Adv. Comput. Sci. Appl.* **7**(8), 164–169 (2016)
4. Baecher, P., Fischlin, M., Gordon, L., Langenberg, R., Lützwow, M., Schröder, D.: Captchas: the good, the bad and the ugly. In: *Sicherheit*, pp. 353–365 (2010)
5. Beheshti, S.M.R.S., Liatsis, P.: Captcha usability and performance, how to measure the usability level of human interactive applications quantitatively and qualitatively? In: *2015 International Conference on Developments of E-Systems Engineering (DeSE)*, pp. 131–136, December 2015. <https://doi.org/10.1109/DeSE.2015.23>
6. Brodić, D., Amelio, A., Ahmad, N., Shahzad, S.K.: Usability analysis of the image and interactive CAPTCHA via prediction of the response time. In: Phon-Amnuaisuk, S., Ang, S.P., Lee, S.Y. (eds.) *MIWAI 2017*. LNCS, vol. 10607, pp. 252–265. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-69456-6\\_21](https://doi.org/10.1007/978-3-319-69456-6_21)
7. Brodić, D., Amelio, A., Draganov, I.R.: Statistical analysis of dice captcha usability. In: *Proceedings of the Information, Communication and Energy Systems and Technologies - 52nd International Scientific Conference, ICEST 2017, Niš, Serbia, June 28–30, 2017*, pp. 139–142 (2017)
8. Brodić, D., Amelio, A., Draganov, I.R., Janković, R.: Exploring the usability of the dice CAPTCHA by advanced statistical analysis. In: Agre, G., van Genabith, J., Declerck, T. (eds.) *AIMSA 2018*. LNCS, vol. 11089, pp. 152–162. Springer, Cham (2018)
9. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995). <https://doi.org/10.1007/BF00994018>

10. DiceCAPTCHA: <http://dice-captcha.com/>
11. Guerar, M., Merlo, A., Migliardi, M.: Completely automated public physical test to tell computers and humans apart: a usability study on mobile devices. *Future Gener. Comput. Syst.* **82**, 617–630 (2018). <https://doi.org/10.1016/j.future.2017.03.012>
12. Iantovics, L.B., Rotar, C., Nechita, E.: A novel robust metric for comparing the intelligence of two cooperative multiagent systems. *Procedia Comput. Sci.* **96**, 637–644 (2016). <https://doi.org/10.1016/j.procs.2016.08.245>. knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 20th International Conference KES-2016
13. Ince, I.F., Salman, Y.B., Yildirim, M.E., Yang, T.: Execution time prediction for 3D interactive captcha by keystroke level model. In: 2009 Fourth International Conference on Computer Sciences and Convergence Information Technology, pp. 1057–1061, November 2009. <https://doi.org/10.1109/ICCIT.2009.105>
14. KentStateUniversity: Pearson's correlation coefficient. <https://libguides.library.kent.edu/SPSS/PearsonCorr>
15. Levenberg, K.: A method for the solution of certain non-linear problems in least squares. *Quart. J. Appl. Math.* **II**(2), 164–168 (1944). <https://doi.org/10.1090/qam/10666>
16. Lupkowski, P., Urbanski, M.: Semcaptchauser-friendly alternative for ocr-based captcha systems. In: 2008 International Multiconference on Computer Science and Information Technology, pp. 325–329, October 2008. <https://doi.org/10.1109/IMCSIT.2008.4747260>
17. Mohamed, M., Gao, S., Sachdeva, N., Saxena, N., Zhang, C., Kumaraguru, P., van Oorschot, P.C.: On the security and usability of dynamic cognitive game captchas. *J. Comput. Secur.* **25**, 205–230 (2017). <https://doi.org/10.3233/JCS-16847>
18. Nguyen, T.T.: Studies of Dynamic Cognitive Game CAPTCHA Usability and Stream Relay Attacks. Doctoral dissertation, California State Polytechnic University, Pomona (2017)
19. Rokach, L., Maimon, O.: *Data Mining With Decision Trees: Theory and Applications*, 2nd edn. World Scientific Publishing Co. Inc, River Edge (2014)
20. Tanikić, D., Marinković, V., Manić, M., Devedžić, G., Randelović, S.: Application of response surface methodology and fuzzy logic based system for determining metal cutting temperature. *Bull. Pol. Acad. Sci. Tech. Sci.* **64**(2), 435–445 (2016). <https://doi.org/10.1515/bpasts-2016-0049>
21. Wikipedia: The captcha test. <http://en.wikipedia.org/wiki/CAPTCHA>
22. Yan, J., El Ahmad, A.S.: Usability of captchas or usability issues in captcha design. In: Proceedings of the 4th Symposium on Usable Privacy and Security, pp. 44–52. SOUPS 2008. ACM, New York (2008). <https://doi.org/10.1145/1408664.1408671>
23. Yu, J., Ma, X., Han, T.: Usability investigation on the localization of text captchas: take chinese characters as a case study. In: Proceedings of the Transdisciplinary Engineering - 24th ISPE Inc., International Conference, vol. 5, pp. 233–242. IOS Press, Singapore (2017)