



An investigation of the usability of image-based CAPTCHAs using PROMETHEE-GAIA method

Sanela Arsić¹ · Djordje Nikolic¹ · Milena Jevtic¹

Received: 29 May 2020 / Revised: 7 September 2020 / Accepted: 7 October 2020 /
Published online: 10 November 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Today, it is difficult to find an adequate Web site with a registration form that is not protected with some automated human proof test. One of the oldest concepts in Artificial Intelligence as a security mechanism based on the Turing Test is CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart). This test was designed to make a difference between the real users and bots and provide security against malicious attacks. The PROMETHEE-GAIA method was employed for ranking different image-based CAPTCHAs according to their usability in this paper. The aim of this study is a comparative analysis of seven image-based CAPTCHAs based on three different criteria: time to find a solution, a number of attempts, and task difficulty. The weights of the considered criteria were calculated objectively by the entropy method, and for the subjective weights, Analytical Hierarchy Process (AHP) was used. A defined research model was applied through four phases. The survey included 320 randomly selected Internet users and experts in computer science who were familiar with CAPTCHA tests. The proposed model suggested which CAPTCHA offered better human accuracy and lower machine attack rates compared to the existing approaches. The obtained results were very helpful to the web administrators because they indicated that this approach could provide useful insights for the decision-makers about the appropriate mechanisms to protect users against cyber-criminal activities and Internet threats.

Keywords Image-based CAPTCHA · Web security mechanism · PROMETHEE-GAIA method · AHP method · Shannon entropy method

1 Introduction

Since the early days of the Internet, different services have been provided to serve a variety of users. Service providers aim to allocate these services and resources for end-

✉ Sanela Arsić
saarsic@tfbor.bg.ac.rs

¹ Engineering Management Department, Technical Faculty in Bor, University of Belgrade, Bor 19210, Serbia

users [8, 41, 53]. Internet users are increasingly encountering the emergence of malware programs and bots (automated scripts) that increase distribution of malicious information, reduce system performance, and waste resources [5, 17]. In order to eliminate these shortcomings, the CAPTCHA test (Completely Automated Public Turing Test to Tell Computers and Humans Apart) was developed in 1950 by Alan Turing [24]. The intention of Artificial Intelligence was to create a machine able to think as a human person. CAPTCHA represents a security mechanism that is used to distinguish human users from malicious computer programs trying to gain illegitimate access to resources [6, 21, 50, 54]. Hence, its main function is to: protect web sites, applications, interfaces, and services such as Google, prevent spam in blogs, and protect email addresses [43].

CAPTCHAs are used for a variety of online applications such as free email accounts, e-commerce, online polling, chat rooms, and many other interactive online services [6, 8]. There are different categories of CAPTCHAs: text-based, audio, video, and image-based [16, 49]. Image-based CAPTCHA represents one of the most effective solutions for machine attacks that the web sites are faced with.

Recognizing objects and images is normal for humans, but for computers it is too complex [25]. The use of this type of CAPTCHA requires minimal interaction with the keyboard, however, it requires significantly increased server processing and the area of a web page [35], but it is widely applied in many societies [35, 51]. In this research several forms of image-based CAPTCHAs were employed. This type of CAPTCHA is connected to the elements of Human-Computer Interaction and it relates to the choice of one or several images from an offered list based on the requested properties whose solving is affected by human factors. Selection of the appropriate CAPTCHAs represent a major challenge for web administrators.

The authors defined an innovative research model that was applied in the paper through four phases. The purpose of this study was to rank seven different image-based CAPTCHAs that were investigated based on their usability by PROMETHEE-GAIA method. The objective of this paper was to investigate the best alternative through two scenarios, in this case, the best ranking CAPTCHA, including the evaluation of three different criteria through the subjective and objective assessment approach. The AHP method was used for the calculation of subjective weights of criteria, and the results were analyzed in scenario 1. Moreover, the Shannon Entropy method was used for the calculation of objective weights in scenario 2.

This study, in addition to the empirical contribution that is reflected in the linking of the scientific literature about CAPTCHAs test and PROMETHEE/GAIA methods which overcome the gap in the scientific literature, also has practical implications in the field of computer science. All web administrators who are dealing with a decision-making dilemma when they want to implement appropriate computer security mechanisms on their websites, can easily make an adequate decision considering a broad variety of indicators of usability using multi-criteria decision-making methods.

The paper is presented in the following sections. Section 2 gives the relevant literature background regarding the introduction and usability of image-based CAPTCHA; Section 3 describes the experimental part and formulation of the four step methodological framework for defining a research topic; Section 4 presents a discussion of the obtained results and analyses the contribution; and, finally, the conclusions and guidelines for the future work are presented in Section 5.

2 Theoretical framework of research

One of the safest and advanced types of CAPTCHA is image-based CAPTCHA. It requires from the users to find and point to the desired image from an image list. Because it is based on image details, it represents an extremely difficult task for a bot to solve [35]. Figure 1 shows an example of image-based CAPTCHA where users have to select all images in which Bill Gates appears [30], and in some cases, users need to select only one image.

In scientific literature, the authors considered different approaches of usability of the CAPTCHA which was one of the fundamental issues relating to it. The usability is related to the process of solving the CAPTCHA test by the Internet users, but if humans cannot use it, then the reason for its application does not exist. The usability also refers to measuring efficiencies of solving the CAPTCHA test through response time, number of attempts, and level of difficulty for the computer user. The usability of image-based CAPTCHAs and text-based CAPTCHAs that verify computer users who join the web was analyzed by [35]. The results of this study showed that the respondents preferred ESP-PIX and ASIRRA, and that statistically significant differences between the CAPTCHAs which were based on task completion time did not exist. Kwon and Cha [30] proposed the technique for analyzing the effectiveness of image annotation based on CAPTCHAs. The results of this technique showed a positive impact on the respondents. Yan and El Ahmad [56, 57] investigated usability aspects of the text-based CAPTCHA design. The results showed that this type of CAPTCHA was difficult for foreigners due to users' language barriers. Table 1 presents the chronological order of overview of the relevant literature concerning the usability of different types of CAPTCHAs (Fig. 2).

Based on the detailed analysis of the previous studies, it could be stated that the application of multi-criteria decision methods in combination with the CAPTCHAs has not been considered. The intention of this paper is to overcome this gap and offer a valuable scientific

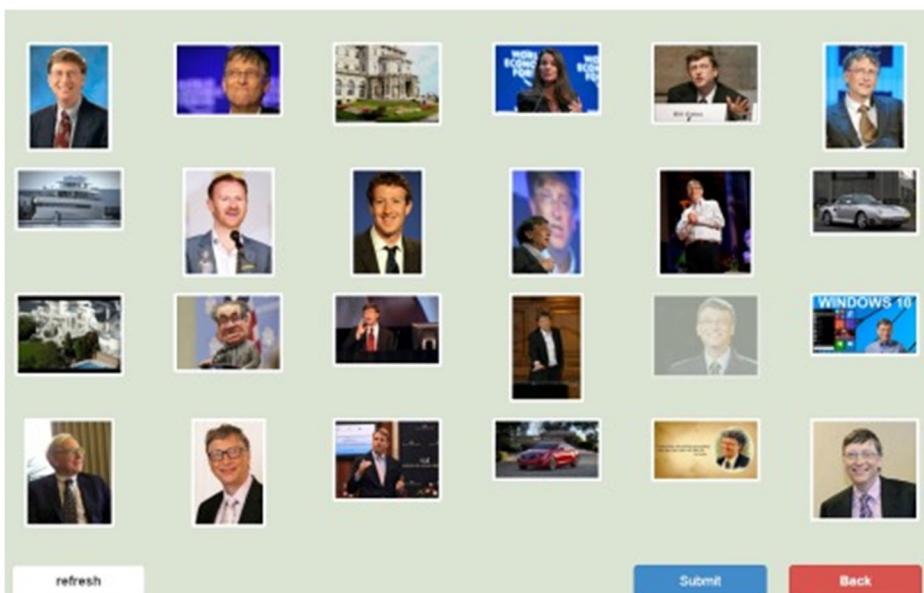


Fig. 1 Example of image-based CAPTCHA Source: [30]

Table 1 Literature review of the usability of CAPTCHA

CAPTCHA type	Year	Author(s)
Image-based	2019	Madathil, K.C., Greenstein, J.S., Horan, K. [35]
	2017	Kwon, S., Cha, S. [30]
	2016	Schryen, G., Wagner, G., Schlegel, A. [47]
	2012	Goswami et al. [22]
	2012	Penninger, S., Meier, S., Federrath, H. [40]
	2010	Madathil, K.C., Alapatt, G.F., Greenstein, J.S. [34]
Text-based	2019	Madathil, K.C., Greenstein, J.S., Horan, K. [35]
	2014	Bursztein et al. [14]
	2013	Shirali-Shahreza et al. [48]
	2012	Pashalidis et al. [39]
	2011	Lee, Y.L., Hsu, C.H. [31]
	2010	Bursztein et al. [15]
Audio-based	2014	Meutzner et al. [36]
	2008	Sauer, G., Hochheiser, H. [45]
	2010	Bursztein et al. [15]
	2013	Shirali-Shahreza et al. [48]
Video-based	2017	Kwon, S., Cha, S. [30]
	2009	Cluever, K.A., Zanibbi, R. [28]
	2008	Cluever, K.A. [27]
Usability of CAPTCHA general	2016	Roshanbin, N., Miller, J. [43]
	2008	Yan, J., El Ahmad, A.S. [57]

contribution by applying PROMETHEE-GAIA methodology which introduces a novel systematic approach to planning and managing the CAPTCHA security mechanism in web pages. In addition, this study also has a practical contribution which is reflected through elimination of the dilemmas that web administrators encounter when they make choices about the application of adequate security mechanisms.

3 Methodology framework for ranking CAPTCHAs

In this paper, the PROMETHEE method (Preference Ranking Organization Method for Enrichment Evaluations) was used for ranking the image-based CAPTCHA according to their usability, from the perspective of the set of Internet users and informatics experts [1, 2, 18]. This method is one of the most prominent MCDM methods which is appropriate for solving very complex decision-making problems [1, 4, 11]. The motive for using the PROMETHEE method for processing the starting data set is in certain advantages of this method over the other MCDM methods. These advantages are reflected in a way of problem structuring, taking into consideration the amount of data that is possible to process, the ability for quantification of qualitative data type, good software support, and presentation of obtained results ([33]; Visual PROMETHEE software, 2012). Analyzing the state of the art of the CAPTCHAs, the authors did not find scientific papers that considered analyzing criteria for ranking CAPTCHAs. Therefore, the significant scientific contribution of this research is reflected in the fact that this methodology can become widely recognized in computer science. The experimental part of the study was conducted through four steps (Fig. 2): (1) problem identification and data collection; (2) determination of criteria; (3) weighting the criteria via AHP and Shannon Entropy methods; and (4) application of the PROMETHEE/GAIA method. The Internet users familiar with CAPTCHA and experts in the field of computer science were included in the study. Respondents evaluated seven

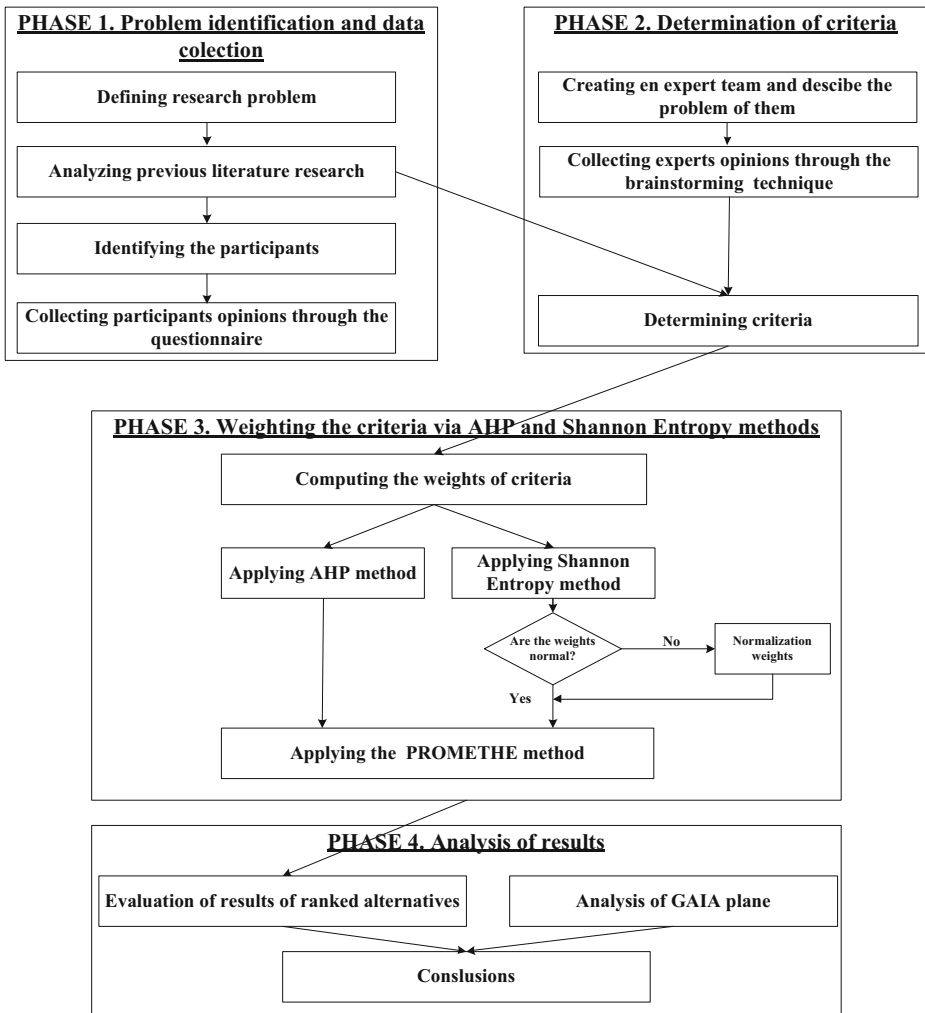


Fig. 2 Proposed research model

different CAPTCHAs image based on three criteria: time to find a solution, a number of attempts, and task difficulty. Two approaches (two scenarios) were used to evaluate the criteria. In scenario 1, the subjective opinion of experts in the field of computer science was examined. The AHP method was used to quantify the data in this scenario. In the second scenario, the objective opinion of the Internet users was quantified with the Shannon entropy method. In the last phase of the research model, the ranking of CAPTCHAs was performed using the PROMETHEE II method, and the GAIA (Geometrical Analysis for Interactive Assistance) plane was used for the graphical representation of the obtained results.

3.1 Data collection

Three hundred and twenty subjects who use Internet daily participated in the experiment. All respondents participated voluntarily and anonymously in research through an online

form. To avoid being influenced, the respondents were not informed about the scope of the analysis, or the collected data types. Data collection was conducted between November and December 2019 and it consisted of two parts. In the first part of the survey, respondents filled the questionnaire with information which related to their demographic's characteristics. Each respondent was characterized by: gender, age, level of education, number of years of the Internet experience, and daily Internet usage represented through the number of hours. The second part of the survey focused on the analysis of image-based CAPTCHAs. The task of each respondent was to find a solution for the following seven image-based CAPTCHAs: Animal in wild (CPT 1), House numbers (CPT 2), Picture of CAPTCHA (CPT 3), Animated character (CPT 4), Face of an old woman (CPT 5), Surprised face (CPT 6), and Worried Face (CPT 7). Three metrics were analyzed to judge the usability of the CAPTCHAs which the respondents rated: time to find a solution, number of attempts, and task difficulty. In addition, for each respondent, the solution time (in seconds) to the CAPTCHAs was measured from the time when the task was started by the respondent until its completion. When it came to the number of attempts, all respondents had between 1 and 3 attempts. A smart phone and tablet were used for the measuring. The criteria - task difficulty was rated on a scale form 1 up to 5 (1 - very easy to solve; 2 - easy to solve; 3 - neutral attitude; 4 - difficult to solve, and 5 - very difficult to solve).

3.2 PROMETHEE and GAIA methods

Among the numerous methods of MCDM, outranking methods show rapid progress due to their flexibility to adapt to the most realistic decision making situations. PROMETHEE method first appeared in the analysis of the efficiency of different services in hospitals by D'Avignon et al. [19] and for comparing different teaching projects by Dujardin [20]. It is now widely used in construction, ecology, agriculture, economy, medicine, computer science, etc., by [3, 4, 9, 10, 42, 55].

The PROMETHEE represents an outranking method, for the final set of alternatives, developed by Brans [12]. This method starts with the formulation of alternatives and a set of criteria, then it is formed as a $m \times n$ decision matrix [46, 58]. In the PROMETHEE method, it is possible to choose one out of six forms of the preference function (Fig. 3) (Usual, U-shape; V-shape; Level; Linear; and Gaussian) where each form could be described with two thresholds (Q and P). The indifference threshold (Q) shows the largest deviation which the decision-maker considers not to be important, while the preference threshold (P) shows the smallest deviation that is crucial for the decision making [13, 23]. Therefore, PROMETHEE method is based on positive (Φ^+) and negative (Φ^-) preference flows for each alternative in the valued outranking relation to rank the alternatives according to the selected preferences (weights) [12]. Positive flow expresses how much the specific alternative dominates over other alternatives, and negative flow expresses how much that alternative is dominated by the others [32, 37, 52, 60].

The success of application of the PROMETHEE method is achieved due to the mathematical properties applied in it. Four major steps are usually involved in making decisions based on PROMETHEE method [11, 12, 38, 52]:

Step 1. Forming a decision-making table. This table includes alternatives and criteria that can be formed based on cardinal (quantitative) or ordinal (quantitative) data.

Shape	Mathematical justification	threshold
	Type I: Usual criterion $P(d) = \begin{cases} 0, d \leq 0 \\ 1, d > 0 \end{cases}$	-
	Type II: Quasi-criterion (U-shape) $P(d) = \begin{cases} 0, d \leq q \\ 1, d > q \end{cases}$	Q
	Type III: Criterion with linear preference (V-shape) $P(d) = \begin{cases} 0, d < 0 \\ d/p, 0 \leq d \leq p \\ 1, d > p \end{cases}$	P
	Type IV: Level criterion $P(d) = \begin{cases} 0, d \leq 0 \\ 1/2, q < d < p \\ 1, d \geq p \end{cases}$	Q,P
	Type V: Criterion with linear preference and indifference area $P(d) = \begin{cases} 0, d \leq q \\ \frac{d-q}{p-q}, q < d \leq p \\ 1, d > p \end{cases}$	Q,P
	Type VI: Gaussian criterion $P(d) = 1 - e^{-\frac{d^2}{2\sigma^2}}$	σ

Fig. 3 List of preference functions [13]

- Step 2. **Calculation of the weight of criteria.** The weight of each criteria is an appropriate index that expresses the priority of criteria toward each other and it indicates the relative importance of that criteria. The greater weight of a criteria implies greater importance compared to other criteria. Weights are non-entering numbers whose sum is equal to 1.
- Step 3. **Evaluation of the preference model.** Based on pairwise comparisons between calculated criteria the deviations are defined. Eq. (1) is used for calculation, where the difference between an evaluated criteria in terms of two *a* and *b* alternatives is represented *d_j*.

$$d_j(a, b) = f_j(a) - f_j(b) \tag{1}$$

Priority indicators are obtained based on all criteria by applying the weight of indicators calculated in the 2nd step W_j , using Eq. (2) to determine the preference of alternatives a and b .

$$\pi(a, b) = \sum_{j=1}^n W_j P_j(a, b); \quad \sum_{j=1}^n W_j = 1 \quad (2)$$

$\pi(a, b)$ is the preference of option a over option b based on the sum of criteria and it implies that option b is prioritized in some criteria compared to option a . $\pi(a, b) \simeq 0$ implies low priority of option a over option b , and $\pi(a, b) = 1$ implies the high priority (complete) of option a over option b [7].

Step 4. Calculation of the preference flow. In order to rank the alternatives Eqs. (3), (4), and (5) are used to calculate the flow of leaving, entering, and net flows, respectively.

$$\Phi^+(a) = \frac{1}{m-1} \sum_{x \in A} \pi(a, x) \quad (3)$$

$$\Phi^-(a) = \frac{1}{m-1} \sum_{x \in A} \pi(x, a) \quad (4)$$

$$\Phi(a) = \Phi^+(a) - \Phi^-(a) \quad (5)$$

$\Phi^+(a)$ is the leaving flow or output of option a (priority of option a over other alternatives), $\Phi^-(a)$ is the entering flow or the input of alternative a (recessive of option a against other alternatives), and $\Phi(a)$ is the net flow of ranking. To facilitate the decision-making process, decision maker usually wants a complete ranking. An alternative with the lowest $\Phi(a)$ has the worst ranking and vice versa [11, 37, 38].

Method for visual representation GAIA complements the PROMETHEE ranking method. GAIA matrix consists of the decomposition of the net outranking flows $\Phi(a)$ [26, 29]. The matrix data are then processed by a Principal Component Analysis (PCA) algorithm, and then displayed on GAIA biplot [52]. The quality of the GAIA representation is given by the delta parameter (Δ), which indicates the quantity of information reflected in the GAIA plane. As a rule of thumb, values larger than 70% can be considered acceptable: little information is lost and the GAIA plane provides a good, reliable representation of the decision problem [32, 38].

3.3 Calculation of the weights of criteria

For solving MCDM problems, an adequate approach for determining the weights of selected indicators is important. Hence, the weights can be classified into two categories depending on the information source: subjective and objective weights [42]. The weight of each criteria is an appropriate index that expresses the priority of criteria toward each other and indicates the relative importance of that criteria.

The weight of criteria was calculated in scenario 1(SC1), through the prism of the subjective weights using Analytical Hierarchy Process (AHP) and the obtained results are shown in Table 3. Three decision-makers who were familiar with the CAPTCHAs were involved in this phase. The AHP method was proposed by Tomas Saaty [44] and its application for determining the weights

of the parameters in MCDA was based on comparing pairs of criteria according to the Saaty scale from 1 to 9, which were then used to calculate the weight of normalized vectors. However, the main problem of the subjective approach in determining the weights of criteria was the inconsistency, which occurred mainly due to the decision-makers inability to always provide a consistent assessment under different schemes to determine the weight and the fact that the process of determining the weight basis depended on the structure of the problem. This problem could be overcome through an objective approach of assigning weights to the criteria, which could be conducted independently of the subjective assessment of the decision-makers and that could be applied when it was not possible to determine reliable subjective weights [33, 46]. This approach was presented in the second scenario.

Scenario 2 (SC2) used objective weights which were calculated based on the Shannon entropy method. This method was useful for measuring the amount of useful information with the data provided [59]. Therefore, the lower the entropy value was, the lower was the degree of disorder of the system, which indicated that if the difference in the value between the evaluated object for the same criteria was high, the criteria would provide more useful information [38]. In order to calculate objective weights, the next four steps were applied [18, 38, 59].

Step 1. The decision matrix was normalized through Eq. (6), in which $f_j(a_i)$ was the evaluated criteria of the a_i option in terms of j criteria.

$$N_{ij} = \frac{f_j(a_i)}{\sum_{i=1}^m f_j(a_i)}; \forall_j \tag{6}$$

Step 2. Matrix E (the entropy) was calculated per index. Equation (7) was used to calculate matrix E .

$$E_{ij} = \sum N_{ij} \times \log(N_{ij}); \forall_j \tag{7}$$

Step 3. The W_j weight matrix was calculated using Eq. (8):

$$W_j = 1 + (k \times E_j); \forall_j \tag{8}$$

where k was constant and it was obtained through Eq. (9), and m was the number of alternatives:

$$k = \frac{1}{\log(m)} \tag{9}$$

Step 4. After the calculation of weights for each criteria, the sum of the weights may not be equal to 1. Thus, the final normalization of weights was performed according to Eq. (10).

$$W_j = \frac{W_j}{\sum W_j}; \forall_j \tag{10}$$

In order to apply the PROMETHEE/GAIA model, besides the determination of the weight of criteria, a set of parameters for each criteria needed to be assigned. These parameters included the impacts of the proposed alternatives, selecting the maximizing or minimizing value, and a preference function with the related thresholds and weight of the criteria.

4 Results and discussion

The results, presented in Table 2, obtained by the analysis of the descriptive statistics of the respondents, indicate that 55.6% men and 44.4% women participated in the study. The respondents were randomly selected, and the only requirement that they had to fulfill was to be Internet user and to be familiar with CAPTCHA. The greatest number of respondents, as many as 43.1%, were older than 41 years. When it came to the level of education of the respondents, 39.1% had MSc diploma and 29.1% had a BSc diploma. Analyzing number of years of Internet experience showed that 58.8% respondents had between 11 and 15 years of experience, and 33.8% had between 6 and 10 years of Internet experience. The interesting fact was that these two groups of respondents were dominant in a sample. On the other hand, the greatest number of respondents, 49.1%, was not spending so much time daily on the Internet (only up to 3 h a day), and more than 12 h a day was chosen by only 2.2% of respondents.

Based on the calculation of the weight of criteria using the AHP method for the SC1, time to find a solution with a dominant value of 0.53 had the highest priority. The next was task difficulty with value 0.26, and criteria number of attempts had the lowest priority with a value of 0.21 (Table 3). Input parameters for SC2 were determined using the Shannon entropy method after collecting data by questionnaire and its results are shown in Table 3 as well. For the SC2, task difficulty with a value of 0.39 had the highest priority, then the number of attempts with value 0.35, and criteria time to find a solution had the lowest priority with a value 0.26.

Thereafter, the obtained weights of each criteria were entered into the Visual PROMETHEE software. All criteria were considered in minimized form (Table 3). Given that most obtained rating in the questionnaire had similar values and minimal difference existed amongst some evaluations, these fluctuations and conflicts could be considered for determining the preference function in which the answer may change with a duplication of evaluations. Therefore, to prevent such occurrence, the domain of changes for each criterion was calculated.

Table 2 Results of demographic characteristics of respondents

Variables	Category	N	Percentage (%)
Gender	Male	178	55.6
	Female	142	44.4
Age	18–30	90	28.1
	31–40	92	28.8
	Over 41 years	138	43.1
	Level of education	High school	61
Level of education	BSc	93	29.1
	MSc	126	39.1
	PhD	40	12.5
	Years of Internet experience	Up to 5	15
6–10		108	33.8
11–15		188	58.8
16–20		9	2.8
Over 21 years		0	0
Daily Internet usage in number of hours	Up to 3 h	157	49.1
	3–6 h	97	30.3
	6–9 h	41	12.8
	9–12 h	18	5.6
	More than 12 h	7	2.2

Table 3 Input parameters of impact matrix

Criteria	Time to find a solution		Number of attempts		Task difficulty	
	Scenario 1 Subjective AHP weights	Scenario 2 Objective entropy weights	Scenario 1 Subjective AHP weights	Scenario 2 Objective entropy weights	Scenario 1 Subjective AHP weights	Scenario 2 Objective entropy weights
Weight min/max	0.53 min	0.26	0.21 min	0.35	0.26 min	0.39
Preference Function	Linear		Level		Level	
Indifference threshold (<i>Q</i>)	0.05		0.05		0.05	
Preference threshold (<i>P</i>)	0.30		0.30		0.30	
<i>CAPTCHAs</i>						
<i>Short</i>						
Animal in wild	CPT 1 2.66		1.14		1.34	
House numbers	CPT 2 2.99		1.10		1.23	
Picture of CAPTCHA	CPT 3 3.47		1.25		1.78	
Animated character	CPT 4 2.79		1.11		1.22	
Face of an old woman	CPT 5 2.52		1.16		1.23	
Surprised face	CPT 6 2.91		1.27		1.62	
Worried Face	CPT 7 4.32		1.26		1.84	

Table 4 Preference flow for the SC1 (Subjective AHP weights)

ALTERNATIVE	$\Phi^+(a)$	$\Phi^-(a)$	$\Phi(a)$	Ranking
CPT 1	0.3047	0.0659	0.2387	4
CPT 2	0.3195	0.0652	0.2544	3
CPT 3	0.0519	0.5094	-0.4575	6
CPT 4	0.3284	0.0165	0.3118	2
CPT 5	0.4349	0.0175	0.4174	1
CPT 6	0.1710	0.1990	-0.0280	5
CPT 7	0.000	0.7369	-0.7369	7

Considering that the data in Table 3 had a quantitative character, a linear function was chosen as the preference function for first criteria (time to find a solution), and level function for other two criteria (number of attempts and task difficulty) with indifference and preference thresholds (q and p) in the 5% and 30% zones, respectively [32]. Estimated impacts and their indices were submitted to PROMETHEE and were given in Table 3. Application of PROMETHEE for the SC1, which was based on subjective AHP weights, and for SC2, which was based on objective entropy weights, produced the following results (presented in Table 4 and Table 5, respectively) for the calculated values of leaving flows $\Phi^+(a)$, entering flows $\Phi^-(a)$, and net flow $\Phi(a)$. All three flows had values between 0 and 1.

The difference in the obtained results can be seen in the ranking of the position of the first three alternatives for both scenarios. The net flow ranking (Table 6) showed that the most preferred alternative for the SC1 was CPT 5 (Face of an old woman), then CPT 4 (Animated character) and CPT 2 (House numbers) followed, respectively. In the SC2 the most preferred alternative was CPT 4 (Animated character), then CPT 2 (House numbers) and CPT 5 (Face of an old woman) followed, respectively. That indicated that for SC1 the most usability image-based CAPTCHA from the perspective of experts was Face of an old woman, while from the prism of the Internet users it was Animated character. There were no difference in the obtained ranking of alternatives for the last four positions in SC1 and SC2. The order of alternatives from the fourth to the seventh position was: CPT 1, CPT 6, CPT 3, and CPT 7, respectively.

In order to determine the scope of preferred relations with given ranking, the analyses of interval of stability for both scenarios (SC1 and SC2) were done and the obtained results were presented in Table 7.

The purpose of the stability interval analysis was to check the robustness of the chosen preference relations. The interval of stability defined the limits within which the range of the weight coefficient of the given criteria could be obtained without influencing the obtained ranking results. A wide stability interval indicated that the ranking did not change even when

Table 5 Preference flow for the SC2 (Objective entropy weights)

ALTERNATIVE	$\Phi^+(a)$	$\Phi^-(a)$	$\Phi(a)$	Ranking
CPT 1	0.5388	0.1619	0.3769	4
CPT 2	0.6058	0.0855	0.5204	2
CPT 3	0.1342	0.7750	-0.6408	6
CPT 4	0.5935	0.0556	0.5379	1
CPT 5	0.5963	0.1458	0.4505	3
CPT 6	0.2167	0.5815	-0.3648	5
CPT 7	0.0000	0.8800	-0.8800	7

Table 6 Complete ranking of CAPTCHA for the selected scenarios

Ranking	Scenario 1		Scenario 2	
	Alternatives	Net-flow Φ	Alternatives	Net-flow Φ
1	CPT 5	0.4174	CPT 4	0.5379
2	CPT 4	0.3118	CPT 2	0.5204
3	CPT 2	0.2544	CPT 5	0.4505
4	CPT 1	0.2387	CPT 1	0.3769
5	CPT 6	-0.0280	CPT 6	-0.3648
6	CPT 3	-0.4575	CPT 3	-0.6408
7	CPT 7	-0.7369	CPT 7	-0.8800

the parameters varied to a wide extent. However, the changes of weight could be done only by one criterion, while relative weights of the other criteria stayed the same. Based on relatively wide stability intervals (Table 7), it can be concluded that the final order of ranking alternatives did not change even when the weight coefficients varied in relatively wide limits.

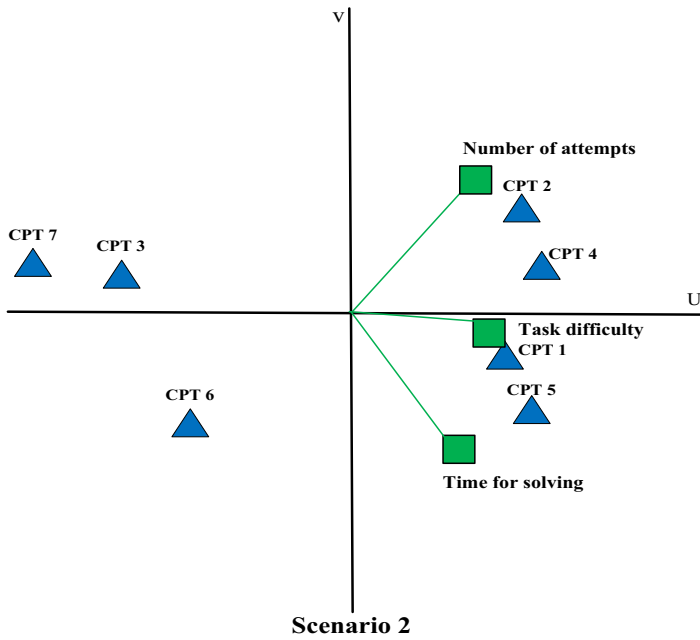
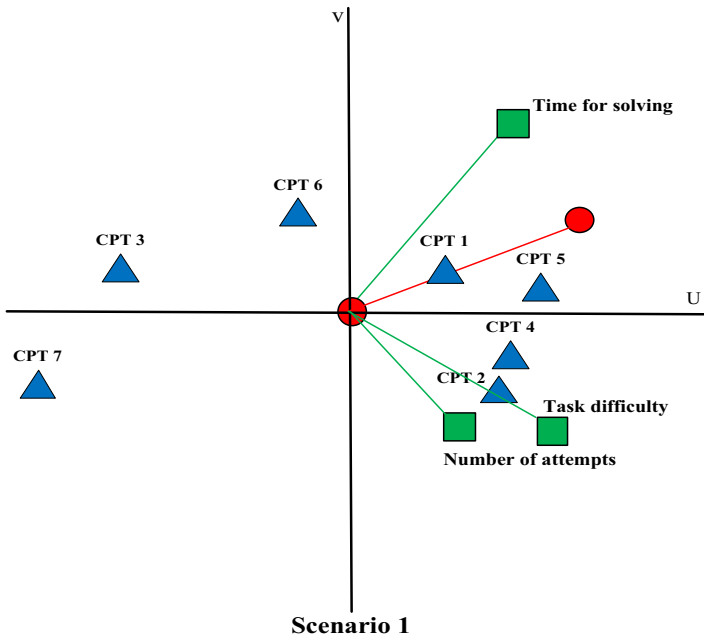
The GAIA biplot provided valuable information in addition to the PROMETHEE ranking. This two-dimensional representation of the problem displayed the relations between alternatives and criteria, indicating strong and weak features of all alternatives and their interaction with the criteria. Alternatives were represented by triangles, and criteria by squares. Based on the position of the criteria in the GAIA plan (Fig. 4), conformity or conflict between individual criteria could be determined. The positions of the alternatives to determine the strength and weaknesses with respect to the criteria could be determined as well. The closer the orientation of the individual criterion axis was, the better the individual alternative was, judging by that criterion. In SC1 according to the criteria number of attempts and task difficulty, the closest position had CPT 2, while according to the criteria time for a solution, the closest position had alternative CPT 1. Moreover, according to the criteria time for a solution, the closest position had alternative CPT 5, according to the criteria number of attempts the closest position had CPT 2, and for criteria task difficulty the closest position had CPT 1 in SC2.

5 Conclusion

Computer science faced numerous challenges when it came to the selection of security mechanisms on the web pages. The most successful technology used as a standard security mechanism was CAPTCHA test. In order to define and implement the most effective measures

Table 7 Weight stability intervals for referent scenarios

Criteria elements	Stability intervals					
	Scenario 1			Scenario 2		
	Weight	Min	Max	Weight	Min	Max
Time to find a solution	0.53	20.84	55.78	0.26	21.28	32.38
Number of attempts	0.21	2.75	53.25	0.35	28.32	41.18
Task difficulty	0.26	23.61	100.00	0.39	21.74	100.00



■ Criteria ▲ Alternatives

Fig. 4 GAIA plane for defined Scenario 1 (Δ - 97.4%) and Scenario 2 (Δ - 99.1%) Source: Visual PROMETHEE software

to overcome the undesirable or malicious Internet bot programs, decision-makers should identify not only the main weaknesses of each web page, but their strong points as well. This would enable the formulation of a comprehensive and coherent set of measures that would neutralize negative effects and support the strengths of each web platform. Preliminary research along these lines requires the application of multi-criteria analysis, considering that the web administrators have to consider a broad variety of indicators of systemic importance. This paper is an example of such an analysis. The application of the PROMETHEE/GAIA method in combination with the entropy model and the AHP method enabled the examination of the usability of CAPTCHA tests. The analysis found that the face of an old woman, based on the opinion of experts in computer science, and Animated character, based on the opinion of the Internet users, were the best tests according to the observed criteria. From the perspective of users, the analysis demonstrated which CAPTCHAs were the easiest to solve, and suggested that these two types of CAPTCHAs were useful because they offered the best human accuracy when solving. The easiness of solving proposed CAPTCHA confirmed that humans could accurately recognize matched faces even under severe distortion, because they were linked to cognitive psychology. Worried face was the worst ranked CAPTCHA according to both groups of respondent. The tested CAPTCHA samples were closely connected to facial expressions, which were difficult for respondents to recognize. But on the other hand, from a web administrator's point of view even more complex CAPTCHAs should be applied to prevent potential machine attacks.

Although the practical and theoretical concepts employed in this study are clearly stated, and the model can be expanded to many other research problems, some constraints are inevitable. One of these constraints can be seen within decision-makers and respondents who were included in the research. Data in this research were collected during a short period of time in order to maintain the compatibility and coherence of the data. Also, the needs and expectations of users are rapidly changing. Therefore, the results of this research are regarded as a snapshot taken in a certain time, and these results can vary as time passes.

Acknowledgments Responsible for the English language in the paper is Sandra Vasković, the English language teacher.

References

1. Abdel-Basset M (2019) A neutrosophic theory-based security approach for fog and mobile-edge computing. *Comput Netw* 157:122–132
2. Abdullah L, Chan W, Afshari A (2019) Application of PROMETHEE method for green supplier selection: a comparative result based on preference functions. *J Ind Eng Int* 15(2):271–285
3. Agha SR, Nofal LG, Nassar HA (2012) Multi-criteria governmental crop planning problem based on an integrated AHP-PROMETHEE approach. *Int J Appl Manag Sci* 4(4):385–406
4. Aherwar A, Singh T, Singh A, Patnaik A, Fekete G (2019) Optimum selection of novel developed implant material using hybrid entropy-PROMETHEE approach. *Mater Werkst* 50(10):1232–1241
5. Alqatani FH, Alsulaiman FA (2020) Is image-based CAPTCHA secure against attacks based on machine learning? An experimental study. *Comput Secur* 88:101635
6. Beheshti SMRS, Liatsis P, Rajarajan M (2017) A CAPTCHA model based on visual psychophysics: using the brain to distinguish between human users and automated computer bots. *Comput Secur* 70:596–617
7. Behzadian M, Kazemzadeh RB, Albadvi A, Aghdasi M (2010) PROMETHEE: a comprehensive literature review on methodologies and applications. *Eur J Oper Res* 200(1):198–215
8. Belk M, Fidas C, Germanakos P, Samaras G (2015) Do human cognitive differences in information processing affect preference and performance of CAPTCHA? *Int J Hum Comput Stud* 84:1–18

9. Ben-Abdallah E, Boukadi K, Lloret J, Hammami M (2019) CROSA: context-aware cloud service ranking approach using online reviews based on sentiment analysis. *Concurr Comput* 1–16
10. Bennani A, Bahi L, Amgaard S (2017) Using a combined approach AHP PROMETHEE to make a decision about roads alternative project: a case study. *Int J Civ Eng Technol* 8(5):856–868
11. Brans J, Mareschal B (2005) PROMETHEE methods. *Multiple Criteria Decision Analysis: State of the Art Surveys*, pp 163–186
12. Brans JP (1982) Lingenierie de la decision, elaboration dinstruments daide a la decision. Methode PROMETHEE. In: Nadeau R, Landry M (eds) *Laide a la decision: nature, instruments et perspectives Davenir*. Presses de Universite Laval, Quebec, pp 183–214
13. Brans JP, Mareschal B, Vincke PH (1984) PROMETHEE: a new family of outranking methods in multi-criteria analysis. In: Brans JP (ed) *Operational research '84*. North-Holland, Amsterdam, pp 477–490
14. Bursztein, E, Moscicki, A, Fabry, C, Bethard, S, Mitchell, JC, Jurafsky, D. (2014). Easy does it: more usable captchas. In: *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, ACM, pp 2637–46
15. Bursztein E, Bethard S, Fabry C, Mitchell J, Jurafsky D (2010) How good are humans at solving captchas? A large scale evaluation. In: *2010 IEEE symposium on security and privacy*, pp 399–413
16. Chandavale AA, Sapkal AM (2015) Systematic approach to measure strength of textbased CAPTCHA. In: *2015 International Conference on Information Processing (ICIP)*, pp 382–387
17. Cheriet M, Bunke H, Hu J, Kimura F, Suen CY (2009) New Frontiers in handwriting recognition. *Pattern Recogn* 42(12):3129–3130
18. Christian AV, Zhang Y, Salifou C (2016) Application of PROMETHEE-GAIA method in the entry mode selection process in international market expansion. *Open J Bus Manag* 4:238–250
19. D'Avignon G, Turcotte M, Beaudry L, Duperré Y (1983) *Degré de specialisation des hopitaux de Quebec*. Technical report. Universite Laval, Quebec
20. Dujardin M (1984) Une evaluation multicritere de projets de remediation a l'echec dans l' enseignement secondaire Belge. In *XIX Meeting of the European Working Group on Multiple Criteria Decision Aid*, Liege, France
21. Fidas CA, Voyiatzis AG, Avouris NM (2011) On the necessity of user-friendly CAPTCHA. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp 2623–2626
22. Goswami G, Singh R, Vatsa M, Powell B, Noore A (2012) Face recognition captcha. In: *Proceedings of the Fifth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, IEEE, pp 412–17
23. Hermgren L, Goonetilleke A, Ayoko GA (2006) Analysis of heavy metals in roaddeposited sediments. *Anal Chim Acta* 571(2):270–278
24. Hidalgo JMG, Alvarez G (2011) Chapter 3 - CAPTCHAs: an artificial intelligence application to web security. *Adv Comput* 83:109–181
25. Karpathy A, Fei-Fei L (2015) Deep visual-semantic alignments for generating image descriptions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 3128–3137
26. Keller HR, Massart DL, Brans JP (1991) Multicriteria decision making: a case study. *Chemom Intell Lab Syst* 11:175–189
27. Kluever KA (2008) *Evaluating the usability and security of a video CAPTCHA*. Rochester Institute of Technology. Thesis. Rochester Institute of Technology. Accessed from <https://scholarworks.rit.edu/theses/163>
28. Kluever KA, Zanibbi R (2009) Balancing usability and security in a video CAPTCHA. In: *Proceedings of the 5th Symposium on Usable Privacy and Security*, ACM, pp 11–14
29. Kocmanova A, Dočekalova M, Lunacek J (2013) PROMETHEE-GAIA method as a support of the decision-making process in evaluating technical facilities. *IFIP Adv Inf Commun Technol* 413:44–53
30. Kwon S, Cha S (2017) CAPTCHA-based image annotation. *Inf Process Lett* 128:27–31
31. Lee YL, Hsu CH (2011) Usability study of text-based CAPTCHAs. *Displays* 32(2):81–86
32. Lopes APF, Muñoz MM, Alarcón-Urbistondo P (2018) Regional tourism competitiveness using the PROMETHEE approach. *Ann Tour Res* 73:1–13
33. Macharis C, Springael J, De Brucker K, Verbeke A (2004) PROMETHEE and AHP: the design of operational synergies in multicriteria analysis. *Strengthening PROMETHEE with ideas of AHP*. *Eur J Oper Res* 153:307–317
34. Madathil KC, Alapatt GF, Greenstein JS (2010) An investigation of the usability of image-based CAPTCHAs. *Proc Hum Factors Ergon Soc Ann Meet* 54(16):1249–1253
35. Madathil KC, Greenstein JS, Horan K (2019) Empirical studies to investigate the usability of text-and image-based CAPTCHAs. *Int J Ind Ergon* 69:200–208

36. Meutzner H, Nguyen V-H, Holz T, Kolossa D (2014) Using automatic speech recognition for attacking acoustic captchas: the tradeoff between usability and security. In: Proceedings of the 30th Annual Computer Security Applications Conference, ACM, pp 276–85
37. Nikolić D, Jovanović I, Mihajlović I, Živković Ž (2009) Multi-criteria ranking of copper concentrates according to their quality – an element of environmental management in the vicinity of copper – smelting complex in Bor, Serbia. *J Environ Manag* 91:509–515
38. Ostovare M, Shahraki MR (2019) Evaluation of hotel websites using the multicriteria analysis of PROMETHEE and GAIA: evidence from the five-star hotels of Mashhad. *Tour Manag Perspect* 30:107–116
39. Pashalidis A, Mavrogiannopoulos N, Ferrer X, Bermejo Olaizola B (2012) For human eyes only: security and usability evaluation. In: Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society, ACM, pp 129–40
40. Penninger S, Meier S, Federrath H (2012) Usability von CAPTCHA-Systemen. In: *GI Sicherheit*, Darmstadt, 195:199–208
41. Plamondon R, Pirlo G, Anquetil É, Rémi C, Teulings H-L, Nakagawa M (2018) Personal digital bodyguards for e-security, e-learning and e-health: a prospective survey. *Pattern Recogn* 81:633–659
42. Radulescu M, Fedajev A, Nikolic D (2017) Ranking of EU national banking systems using multi-criteria analysis in the light of BREXIT. *Acta Oeconomica* 67(4):473–509
43. Roshanbin N, Miller J (2016) ADAMAS: interweaving unicode and color to enhance CAPTCHA security. *Futur Gener Comput Syst* 55:289–310
44. Saaty TL (1980) *The analytic Hierarchy process*. McGraw-Hill, New York
45. Sauer G, Hochheiser H (2008) Towards a universally usable CAPTCHA. In: Proceedings of the 4th symposium on usable privacy and security, pp 2–5
46. Savic M, Nikolic D, Mihajlovic I, Zivkovic Z, Bojanov B, Djordjevic P (2015) Multi-criteria decision support system for optimal blending process in zinc production. *Mineral Process Ext Metall Rev Int J* 36(4): 267–280
47. Schryen G, Wagner G, Schlegel A (2016) Development of two novel face-recognition CAPTCHAs: a security and usability study. *Comput Secur* 60:95–116
48. Shirali-Shahreza S, Penn G, Balakrishnan R, Ganjali Y (2013) Seesay and hearsay captcha for mobile interaction. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, pp 2147–56
49. Starostenko O, Cruz-Perez C, Uceda-Ponga F, Alarcon-Aquino V (2015) Breaking text-based CAPTCHAs with variable word and character orientation. *Pattern Recogn* 48(4):1101–1112
50. Thomas AO, Rusu A, Govindaraju V (2009) Synthetic handwritten CAPTCHAs. *Pattern Recogn* 42(12): 3365–3373
51. Trukenbrod AK, Backhaus N, Thomaschke R (2020) Measuring subjectively experienced time in usability and user experience testing scenarios. *Int J Hum Comput Stud* 138:102399
52. Vego G, Kučar-Dragičević S, Koprivanac N (2008) Application of multi-criteria decision-making on strategic municipal solid waste management in Dalmatia, Croatia. *Waste Manag* 28:2192–2201
53. Vikram S, Fan Y, Gu G (2011) Semage: a new image-based two-factor CAPTCHA. In: Proceedings of the 27th Annual Computer Security Applications Conference, ACM, pp 237–246
54. von Ahn LV, Blum M, Hopper NJ, Langford J (2003) CAPTCHA: Using hard AI problems for security. *EUROCRYPT'03: Proceedings of the 22nd international conference on theory and applications of cryptographic techniques*, pp 294–311. Accessed 13 Jan 2020
55. Wang JJ, Wei CM, Yang D (2006) Decision method for vendor selection based on AHP/PROMETHEE/GAIA. *J Dalian Univ Technol* 46(6):926–931
56. Yan J, El Ahmad AS (2008a) Usability of CAPTCHAs or usability issues in CAPTCHA design. Symposium on usable privacy and security, Pittsburgh
57. Yan J, El Ahmad AS (2008b) Usability of CAPTCHAs or usability issues in CAPTCHA design. In: Proceedings of the 4th Symposium on Usable Privacy and Security, Pittsburgh, Pennsylvania, USA, pp 44–52
58. Yu X, Xu Z, Ma Y (2013) Prioritized multi-criteria decision making based on the idea of PROMETHEE. *Proc Comput Sci* 17:449–456
59. Zhi-Hong Z, Yi Y, Jing-nan S (2006) Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment. *J Environ Sci* 18:1020–1023
60. Zindani D, Kumar K (2018) Material selection for turbine seal strips using PROMETHEE-GAIA method. *Materials Today: Proceedings* 5:17533–17539