An Integrated Model for Evaluation of Big Data Interactive Systems Under Uncertainty

Adeleh Asemi
University of Malaya

Asefeh Asemi (asemi.asefeh@uni-corvinus.hu)
Corvinus University of Budapest https://orcid.org/0000-0003-1667-4408

Research

Keywords: Usability Evaluation, Usability testing, Big data interactive systems, Fuzzy Multi Factors Evaluation (MFE), Fuzzy Inference System (FIS), Experts' Judgment.

Posted Date: September 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-915588/v1

An integrated model for evaluation of big data interactive systems under uncertainty

Adeleh Asemi
Department of Software Engineering, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia, email: adeleh@um.edu.my
Asefeh Asemi (Corresponding Author)
Doctoral School of Economic, Business, & Informatics, Corvinus University of Budapest, 1093 Budapest, Fovam ter 8., Hungary, email: asemi.asefeh@uni-corvinus.hu

Abstract— The study aimed to propose a judgment-based evaluation model for usability evaluating of big data interactive system s. Human judgment is associated with uncertainty and gray information. We used the fuzzy technique for integration, summarization, and distance calculation of quality value judgment. The proposed model is an integrated fuzzy Multi Factors Evaluation (MFE) model based on experts' judgments in HCI, ISPD, and AMLMs. We provided a Fuzzy Inference System (FIS) for scoring usability evaluation metrics in different big data interactive system s. A multi-model big data interactive system is implemented for experimental testing of the model. The achieved results from the proposed model and experimental tests are compared using statistical correlation tests. The results show the ability of the proposed model for usability evaluation of big data interactive systems without the need for conducting empirical tests. It is concluded that applying a dataset in a neuro-FIS and training system cause to produce more than a hundred effective rules. The findings indicate that the proposed model can be applied for big data interactive system evaluation, informative evaluation, and complex empirical tests. Future studies may improve the FIS with the integration of artificial neural networks.

Keywords— Usability Evaluation, Usability testing, Big data interactive systems, Fuzzy Multi Factors Evaluation (MFE), Fuzzy Inference System (FIS), Experts' Judgment.

1 INTRODUCTION

The COVID-19 scenario has led to a significant expansion in using interactive systems. Therefore, evaluating and comparing big data interactive systems has received more and more attention from researchers. The most popular way for usability evaluation in big data interactive systems is conducting empirical tests. However, sometimes the empirical test is very tough and expensive due to securing space, developing the tests, hiring participants, etc. Moreover, in prototype testing, we cannot apply empirical testing. Evaluating big data interactive systems is extensively investigated in Human-Computer Interaction (HCI) in interdisciplinary fields. It is an activity that examines the degree to which an interactive system satisfies user goals and expectations [1]. Various studies consider evaluating interactive systems. An et al. (2017) proposed a network data envelopment analysis model to measure the interactive relationship between system components [2]. Their model evaluates a parallel system with two interactive components in only two centralized and non-centralized modes. This model specifically evaluates networks with interactive components. However, developing a multi-mode system for evaluation is a proper method for comparing all of the interactive systems, which we adopt from this study. Benson & Powell (2015) propose an interview protocol to improve investigative interviewing of children in training interactive systems [3]. Various studies empirically evaluate the usability of HCI interactive systems through System Usability Scale [4], usability heuristics [5], or multiple criteria [6]; [7]; [8]; [9]. While the type of interactive system is not considered to select proper criteria for evaluation. For mobile applications, educational systems, and agriculture systems, they used the same heuristics or scales with the same level of importance for all criteria. The basic view in evaluating interactive systems is empiricism [10]. Thus, these kinds of evaluations of

interactive systems are dominant. Empirical evaluations pay to the user's needs, and it requires careful planning in method selection. An empirical evaluation is essential to attend to the users' behavior and their interaction with the system. There are several methods in empirical evaluation [3]; [11]; [12] as well as in books with more specific framing about the empirical usability evaluation and the user experience [1]. These methods include observation, interview, focus groups, user testing, field testing, field studies, questionnaires, surveys, diary studies, and empirical usability testing. Experimental usability testing is a summative assessment that often occurs late in the design phase. It is two types of evaluations, including formative and summative. Formative evaluation focuses on usability problems, and summative evaluation evaluates the effectiveness of the final design [13]. So, developers are looking at methods that can be used earlier when only an immature design is available [10]. The empirical usability testing should provide objective data that is difficult and expensive. It costs money and time to set up and execute a good empirical study. Costs revolve around securing space, the development time of the tests, hiring participants, etc. [14]. Therefore, we only focus on the formative evaluating the interactive systems. Formative evaluations focus on identifying usability problems through a redesign. Some of the most common expert-based usability assessment methods include reviewing guides based on interaction design guidelines, innovative assessment, cognitive enhancements, usage paths, formal usability inspections, and innovative marches [15]. Recently, these experts-based techniques have become extended through proposing new models of evaluation factors [16]; [17], new heuristics [18]; [19], usability evaluation in a new generation of software [20]; [21], integration of expert-based techniques and using machine learning techniques to predictive several usability factors based on expert' opinion [22]; [55]. As we discussed, the expert-based usability evaluation in the literature is completely static. To our best knowledge, there is not any usability evaluation method that comprehensively and dynamically evaluates a big data interactive system with considering multiple factors, uncertainty, and experts' judgments. The past researchers used artificial intelligence and active learning methods only for measuring or predicting factors and not obtaining the effect of factors dynamically. For example, Imel et al. (2019) proposed a system based on machine learning techniques to predicting and generating feedback in a usability therapy assessment [22]. Our objective is to propose a new model to evaluate big data interactive systems that covers all the needs of formative and non-empirical evaluation. This model dynamically determines the effect of evaluation factors based on the big data interactive system using a Fuzzy Inference System (FIS). Therefore, big data interactive systems will have a different and proper formulation for their usability evaluation. We consider ISO standards and the most popular usability evaluation factors presented in the literature to be used in FIS. Finally, we apply the fuzzy Multiple Factors Evaluation (MFE) approach for i) comparing two or more big data interactive systems or ii) evaluating an interactive system individually. In fuzzy MFE and FIS, we fuzzify human opinion through designing proper fuzzy Membership Functions (MFs). Fuzzification of expert judgment provides a mapping of the human decision to crisp numbers [23]; [24]. We implement a multi-model Big data interactive systems for People with Disabilities (ISPDs) based on four active machine-learning methods. This multi-model system will be evaluated using the proposed evaluation model for all four applied machine-learning methods. In this study, the methodology of Active Machine Learning Methods (AMLMs) evaluation using the fuzzy MFE method is explained in section 2. In continue, section 3 provides the results of consistency, assessment of AMLMS effects, and evaluating fuzzy MFE method. Section 4 concludes the study.

2 Related Works

Based on our objective, the study is limited to formative evaluation of interactive systems, dynamically, and based on multiple factors, expert judgments, and uncertainty. In this section, we discuss the most related online published works in 1970-2020 and, present the difference between associated works and the current study. Usability evaluation is a multi-criteria decision-making problem that involves multiple fuzzy factors. The MFE methods address the uncertainty and user preferences also can be applied for formative usability evaluation. Fuzzy distance calculation and pairwise comparison are the most popular methods used in fuzzy MFE. These methods are used to conduct the usability evaluation of different systems [25]; [26]; [27]; [28]. Ramanayaka, Chen & Shi (2019) have applied MCDM methods, for the weighting of usability factors to reveal the level of each factors' contribution to the usability index of library websites [28]. Chang and Dillon (2006) for the first time, used fuzzy set theory in usability evalua-

tion [29]; [30].. They defined six dimensions for usability evaluation as System feedback, Consistency, Error prevention, Performance/ efficiency, User like/ dislike, and Error recovery. Chang and Dillon (2006) evaluated their FIS in several different user interfaces [29]. Kumar, Tadayoni, & Sorensen (2015) defined five fuzzy usability attributes as navigation, presentation, learnability, customizing, and task support [31]. Huddy et al. (2019) suggest a consolidated, hierarchical usability model with a detailed taxonomy for specifying and identifying the quality components and measuring usability [32]. The studies as mentioned earlier solve the uncertainty of quality components and address the user preferences in usability evaluations. They have proposed FISs for a dynamic usability evaluation. This FIS is provided for general evaluation [29]; [30]; [32] or a specific application [31]. In these studies, the evaluation metrics are the input of FIS, and the output is the score of usability. Therefore, in provided FIS, when the fuzzy value of evaluation metrics changes, the system gives a new score for usability. However, in this study, when the interactive system changes, the system generates a new formula for usability evaluation. In the new formula, the effect of usability factors (variables coefficient) takes a new value. The proposed FIS provides the effect of evaluation metrics (outputs of FIS) based on types of big data interactive systems (inputs of FIS).

3 METHODOLOGIES

This research has been performed both experimentally and analytically. We propose a model for evaluating computer big data interactive systems based on experts' judgments. This model resolves to evaluate big data interactive systems in conflict problems. It is proper for situations that conducting an empirical evaluation of an interactive system is costly or complicated. The proposed model has three phases (Figure 1). In the first phase, an expert system is implemented to formula usability evaluation for big data interactive systems, dynamically. In the second and third phases, the big data interactive systems evaluate or compare accordingly based on the generated formula in the first phase. If a practitioner wants to evaluate an interactive system individually, they need to apply the second and third phases. Also, if a practitioner wants to compare two big data interactive systems, then they need to use the first and third phases.

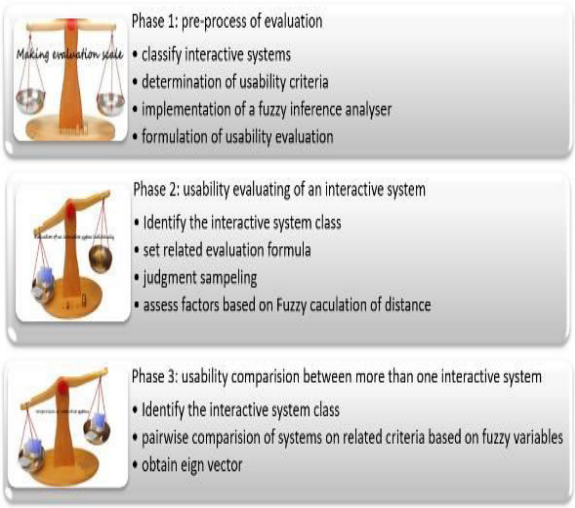


Fig. 1 Proposed Model

3.1 Pre-processing of Model

In the first phase, a FIS is presented, which uses the interactive system as an input and produces the effect of usability criteria as an output. We used the effects received in the formula of usability evaluation as the new variable coefficients. Therefore, we obtain a new formulation of usability evaluation, which is suitable for evaluating the specified interactive system.

3.1.1 Classify Big data interactive systems

Types of big data interactive systems are determined through the classification of big data interactive

systems. In the cause of the variety of big data interactive systems, we consider the four main HCI classifications i) human contribution, ii) human activities, iii) system objective and iv) information processing. The first classification is based on the level of human contribution. That is adopted from Sheridan & Verplank (1978), which proposed groups of human contributions for big data interactive systems [33]. Figure 2 shows a 10-point scale of groups of the human contribution that is provided in an interactive system.

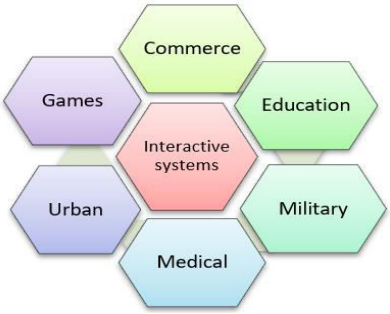


Fig. 2 Groups of Human Contribution in Big data interactive systems

In implementating our expert system, these10 levels form 10 inputs of the system. The classification of human contribution is one of the factors to determine interactive system type. The second classification is associated with user actions. These actions include instruction, conversation, manipulation & navigation, and exploration [34]. It is important to note that these actions can do together. Different methods are used in user interface development. The first method is to allow the user to issue instructions to the system while performing tasks. The second method can be based on the user's conversation with the system. In the third method, the user can manipulate an environment of virtual objects and go their own way. The fourth method is based on a structured information presentation system. This system allows users to learn things without having to ask specific questions. The third classification is based on the purpose of the interactive system. Usability evaluation is directly affected by the purpose of the interactive system. An interactive learning system and an interactive medical system have different users, and the users have different needs. The most proper system with a high usability level is the system, which has the maximum mapping to user needs. We conducted a survey on 182 articles that is resulted from the search in Web of Science Core Collection, in 2018-2019 with “interactive system” search key and in the category of computer science. Component factor analysis in “IBM SPSS statistics 25” is used to classify purposes of interactive systems in these articles. Finally, we obtained six classes for purpose of interactive systems and named business, games, urban, education, medical, and military (Figure 3). The classes have different popularities in collected articles. For example, commerce has the most popular and the military has less popularity among interactive systems.

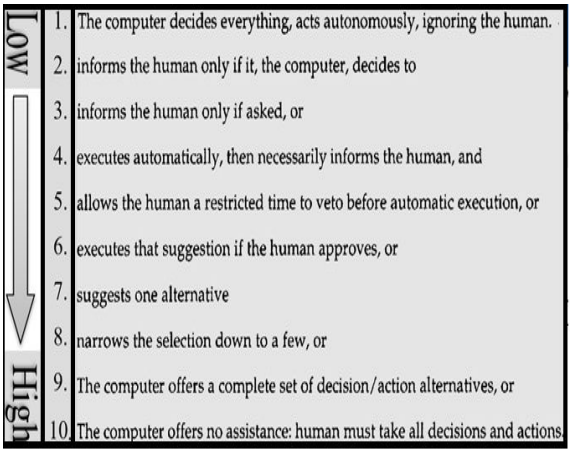


Fig. 3 Interactive Systems Purposes

Interactive systems with the purpose of commerce personalize electronic commerce environments based on Human Factors. Today, personalization is everywhere. Interactive systems with education purposes include e-learning, virtual learning, learning management, and learning services like producing data sets or exploring databases [35]. Interactive systems with medical purposes consist of medical decision support systems, therapist robots, surgery robots, health information systems, and therapeutic systems in interaction with the physician, patient, or expert. The final classification is based on the level of information processing. We adopted a four-level of information processing [36] (Figure 4). In this four-level model, almost all the components of human information processing are obtained during information processing by cognitive psychologists. The performance of different levels in processing operations overlaps in time. Levels can also be coordinated in "perception-action" cycles to provide a precise serial sequence from stimulus to response.

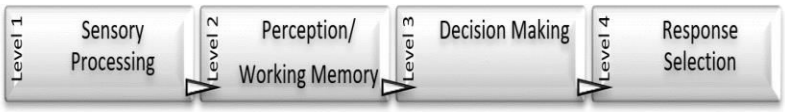


Fig. 4 Levels of Information Processing in Interactive Systems

The first level includes the positioning and orienting of sensory receptors, sensory processing, initial preprocessing of data before complete comprehension, and selective attention. The second level involves the conscious understanding and manipulating information processed and retrieved in working memory. The third level is where decisions are made based on cognitive processing. The last level, the fourth level, involves executing a response or action consistent with the choice of decision.

3.1.2 Determining Usability Criteria

In this section, we focused on two sources of usability evaluation criteria: i) ISO standards, and ii) literature review. According to ISO 9126, the usability feature is defined as "the ability of a software product to be understood, learned, used and attractive to the user when used in specific circumstances" (ISO 9126-2 2001). ISO 9241-11 defines usability as "the extent to which a product is used by specified users to achieve specified goals with specific effectiveness, efficiency, and satisfaction." The definitions of effectiveness, efficiency, and satisfaction are similar in ISO 9241 and ISO 9126, except that ISO 9126 is software-based, and ISO 9241-11 is user-based. The latest revision of 9241-11 proposes eight criteria for interactive systems usability (learnability, regular use, error protection, accessibility, maintainability, effectiveness, efficiency, and satisfaction) [37]. Error protection is minimizing the possibility that users can make errors that could lead to undesirable consequences. In the current research, from these eight criteria, we exclude regular use, maintainability, and satisfaction because we are focusing on a formative evaluation. There is not a ready software product to use users' opinions for the informative evaluation. The usability predicts based on the expert opinion. Therefore, the regular use, maintainability, and satisfaction metrics are not measurable in this stage. In literature, researchers consider a wide range of criteria for usability evaluation. We selected the high cited articles with more than 100 citations in google scholar that provided a list of usability metrics (Figure 5). Sharp et al. (2019) is the most associated and latest work that proposes six criteria (effectiveness, efficiency, safety, utility, learnability, memorability) for usability evaluation, especially in interactive systems [34]. The system utility is directly associated with how its performance is appropriate based on the users' needs. One of the users' needs is the simplicity of using a system. Ease of learning methods is essential to use a system. Users do not like to spend a lot of time learning how to work the system. This problem is especially important for interactive products intended for daily usage. However, many users find this tedious, complex, and time-consuming. It seems that if most users are not able to spend their time learning using the system, it is necessary to create a wide range of learning capabilities for the system. Since we are going to conduct a formative evaluation, memorability, and safety. Sharp et al. (2019) define memorability as "when users return to the design after a period of not using it, how easily can they reestablish proficiency [34]." In the current study, we do not consider flexibility and security evaluation, and we only focus on usability, so we exclude the safety factor as well.

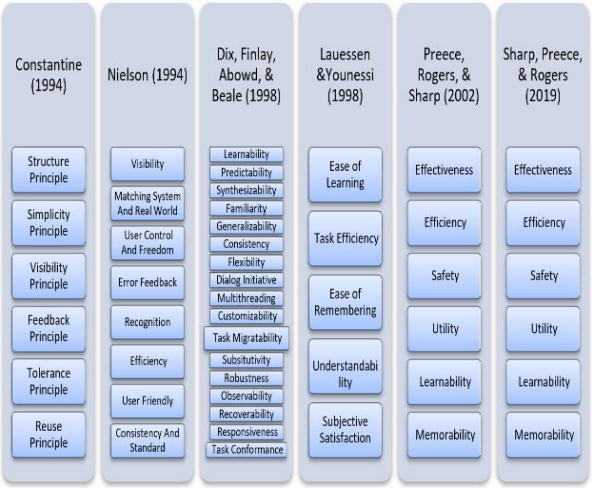


Fig. 5 Usability Evaluation Metrics in High Cited Studies

In this study, we adopt the expert measurable and formative usability evaluation principles collected from standards and literature (Figure 6). Therefore, the final usability evaluation metrics that we select for the output of the usability evaluation formula are effectiveness, efficiency, error protection, learnability, and utility. Accessibility and utility have overlap definitions so, we combine them under utility.

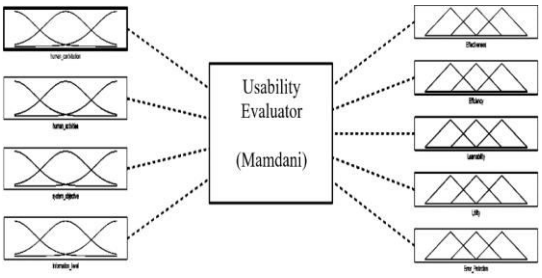


Fig. 6 Selection of Usability Evaluation Metrics for Big data interactive systems

3.1.3 Implementing a Fuzzy Inference Analyzer

Usability criteria are generally in two groups fuzzy variables and linguistic variables. The interactive system is determined based on four criteria. There are three criteria of fuzzy variables. They include user participation, user activity, and information processing. In this study, FIS was designed for usability evaluation using MATLAB with a fuzzy logic toolbox. In this study, we implemented a Mamdani-based FIS. This system was designed to measure the influence of the usability criteria on the whole interactive system (Figure 7). In this method, a fuzzy control strategy is used to plot the given inputs through rules, and produce an output based on these rules. The input is four classifications of big data interactive systems, and the output is five usability metrics. One of the inputs (System Aim) is not considered as a fuzzy variable because we only determine one main objective for an interactive system. However, the other three inputs and five outputs are fuzzy variables.

Fig. 7 Usability Evaluator FIS with Four Inputs and Five Outputs

The designed system is based on fuzzy MFs and if-then rules. The MFs and generated rules help to fuzzy and eliminate fuzzy variables, which is called fuzzification. In fuzzification, perform the process of converting a fuzzy output to a clear output in FIS. The input for the FIS is a fuzzy set, and the output is a single number. An MF is a curve with membership rates between 0 and 1. The MF represents a fuzzy set and is usually denoted by μA. In the fuzzy set, for an element x of X, the value of μA is called the membership degree x. Membership degree, μA (x) determines a degree of membership of the element x in the fuzzy set. A value of 0 shows that x is not a member of the fuzzy set. A value of 1 show that x is a full member of the fuzzy set. Specifies values between 0 and 1 indicate the fuzzy members. Fuzzy logic has eleven internal MFs, and these functions are made up of several essential functions, including linear fragment functions, Gaussian distribution function, sigmoid curves, and quadratic & cube polynomial curves. We determine the MFs for interactive system type inputs and usability metrics output according to the suitability of MF in representing fuzzy variables [38]. Figure 8 shows the designed MFs of inputs of big data interactive systems classes. Human contribution MFs have ten trapezoidal MFs representing the ten groups of human contribution (Figure 8.a). Human activities have 4 Gaussian MFs representing instructing, conversing, manipulating/ navigating, and exploring/ browsing (Figure 8.b). The simplest MFs are formed using straight lines, and the simplest is a triangular MF that we used for crisp input (system purpose) (Figure 8.c). We defined four polynomial curves representing the information processing levels since each level includes the lower levels (Figure 8.d).

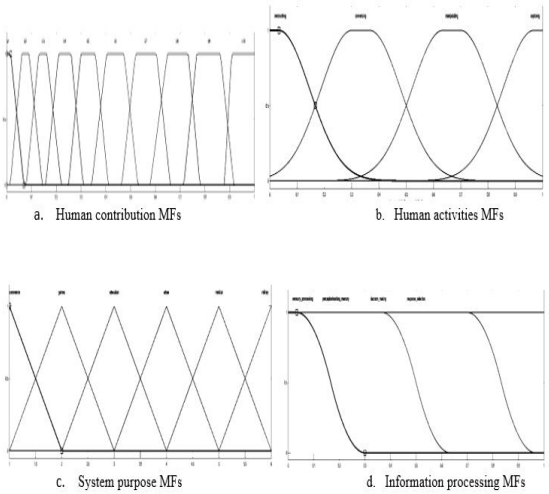


Fig. 8 Input Variables MFs

The output variables have the same MFs. A nine fuzzy scale of importance representing nine Gaussian curves was applied to determine the importance degree for each usability metric (Figure 9).

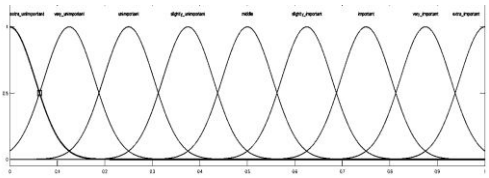


Fig 9. Usability Metrics MFs

Finally, we design if-then rules in the relation between interactive system types, and the effect of usability metrics to predict the effect of usability metrics using fuzzy inferencing (Figure 10).

1. If (human_contribution is L1) then (Learnability is extra_unimportant) (1)
2. If (human_contribution is L10) then (Learnability is extra_important) (1)
3. If (human_contribution is L5) then (Learnability is middle) (1)
4. If (human_contribution is L5) and (human_activities is conversing) then (Learnability is middle) (1)
5. If (human_contribution is L1) and (human_activities is instructing) then (Learnability is very_unimportant) (1)
6. If (human_contribution is L3) and (human_activities is conversing) and (information_level is response_selection) then (Learnability is extra_important)(Utility is slightly_important)(Error_Protrction is very_important) (1)
7. If (system_objective is medical) then (Effectiveness is extra_important)(Learnability is important)(Error_Protrction is extra_important) (1)
8. If (human_contribution is L6) and (human_activities is conversing) and (system_objective is education) and (information_level is decision_making) then (Effectiveness is important)(Efficiency is important)(Learnability is extra_important)(Utility is extra_important)(Error_Protrction is slightly_important) (1)
9. If (human_contribution is L2) and (human_activities is instructing) and (system_objective is military) and (information_level is response_selection) then (Effectiveness is extra_important)(Efficiency is extra_important)(Learnability is very_unimportant)(Utility is middle)(Error_Protrction is extra_important) (1)
10. If (human_contribution is L3) and (human_activities is instructing) and (system_objective is urban) and (information_level is response_selection) then (Effectiveness is important)(Efficiency is slightly_important)(Learnability is extra_important)(Utility is extra_important)(Error_Protrction is middle) (1)
11. If (human_contribution is L4) and (human_activities is instructing) and (system_objective is commerce) and (information_level is decision_making) then (Effectiveness is very_important)(Efficiency is important)(Learnability is extra_important)(Utility is very_important)(Error_Protrction is important) (1)
12. If (human_activities is conversing) and (system_objective is games) and (information_level is decision_making) then (Effectiveness is important)(Efficiency is extra_important)(Learnability is important)(Utility is important)(Error_Protrction is middle) (1)
13. If (human_contribution is L5) and (human_activities is conversing) and (system_objective is urban) and (information_level is decision_making) then (Effectiveness is extra_important)(Efficiency is important)(Learnability is extra_important)(Utility is extra_important)(Error_Protrction is very_important) (1)
14. If (human_contribution is L7) and (human_activities is manipulating) and (system_objective is commerce) and (information_level is perception/working_memory) then (Effectiveness is slightly_important)(Efficiency is slightly_important)(Learnability is slightly_important)(Utility is extra_important)(Error_Protrction is slightly_important) (1)
15. If (human_contribution is L8) and (human_activities is exploring) and (system_objective is medical) and (information_level is sensory_processing) then (Effectiveness is extra_important)(Efficiency is very_important)(Learnability is middle)(Utility is middle)(Error_Protrction is slightly_important) (1)
16. If (system_objective is military) then (Effectiveness is extra_important)(Efficiency is extra_important)(Learnability is unimportant)(Utility is important)(Error_Protrction is extra_important) (1)
17. If (system_objective is games) then (Effectiveness is important)(Efficiency is very_important)(Learnability is very_important)(Utility is very_important)(Error_Protrction is slightly_unimportant) (1)
18. If (system_objective is education) then (Effectiveness is slightly_important)(Efficiency is slightly_important)(Learnability is extra_important)(Utility is extra_important)(Error_Protrction is slightly_unimportant) (1)
19. If (system_objective is commerce) then (Effectiveness is very_important)(Efficiency is important)(Learnability is extra_important)(Utility is extra_important)(Error_Protrction is slightly_unimportant) (1)
20. If (human_contribution is L1) and (human_activities is instructing) and (system_objective is urban) and (information_level is sensory_processing) then (Effectiveness is important)(Efficiency is extra_important)(Learnability is very_unimportant)(Utility is very_important)(Error_Protrction is very_important) (1)
21. If (human_contribution is L4) and (human_activities is conversing) and (system_objective is education) and (information_level is perception/working_memory) then (Effectiveness is important)(Efficiency is slightly_important)(Learnability is very_important)(Utility is very_important)(Error_Protrction is unimportant) (1)
22. If (information_level is response_selection) then (Effectiveness is very_important)(Efficiency is very_important)(Learnability is very_unimportant)(Utility is very_important)(Error_Protrction is extra_important) (1)
23. If (information_level is sensory_processing) then (Effectiveness is very_important)(Efficiency is very_important)(Learnability is extra_important)(Utility is important)(Error_Protrction is unimportant) (1)
24. If (human_contribution is L9) and (human_activities is exploring) and (system_objective is urban) and (information_level is decision_making) then (Effectiveness is very_important)(Efficiency is very_important)(Learnability is very_important)(Utility is very_important)(Error_Protrction is important) (1)

Fig. 10 A Part of if-then Rules in the FIS

3.1.4 Formulating usability evaluation

This study aimed to generate evaluation scores for all types of big data interactive systems. We proposed a fuzzy approach to creating the effect of usability metrics. We make a usability evaluation formula (Equation 1) based on the relationship between effects obtained and the final evaluation score, to calculate the final evaluation score.

$$Usability = \sum_{i=1}^{5} \frac{\sqrt{MV_i * MI_i^2}}{RI} \qquad (1)$$

Usability is a variable that keeps the final score of usability calculated in this equation (1). Variable "i" is a counter that counts from 1 to 5 since we have five usability metrics. Here, usability is a standard value. Arithmetic means to divide into random usability indices. Usually, based on the simulation method and the number of matrices, calculate different RIs. We adopt RI from Noble and Sanchez (1993) [39], which carried out 2500, 1000, and 5000 simulation runs. Random usability index (RI) is associated with previous usability experience of similar big data interactive systems (Table 1). If a similar system is very successful in providing usability, then the usability RI is 0.4, and if we do not have experience in a similar case, then the RI is 1.0.

TABLE 1 Random Usability Index Table Based on the Noble & Sanchez (1993) Simulation [39]

	Very successful	Successful	No experience	Unsuccessful	Very unsuccessful
RI	0.4	0.8	1.0	1.1	1.2

Variable MI is the effect of usability metrics generated by a FIS. MI1, MI2, MI3, MI4, MI5 are the effects of effectiveness, efficiency, learnability, utility, and error protection correspondingly. Variable MV is the value of usability metrics, which is calculated through MFE methods. MV1, MV2, MV3, MV4, MV5 are the values of effectiveness, efficiency, learnability, utility, and error protection. In the second and third phases, we explained that how we can obtain MV, when we have one interactive system for evaluation, or when we have multiple big data interactive systems for comparison.

3.2 Usability Evaluating of an Interactive System

In the second phase, the following steps should be conducted for evaluating an interactive system. A distance-based multi-factor evaluation method is proposed to provide the value of metrics (MV) in the usability formula.

3.2.1 Identify the Interactive System Class

This section is a typical section for phases 2 and 3. We need to determine the state of the entire interactive system in each class that is an input of the fuzzy system. For example, in an interactive system, the human contribution is at the sixth level, human activities are at the conversing level, and the purpose of the system is education. Also, the information processing contains all four steps of processing (Figure 11).

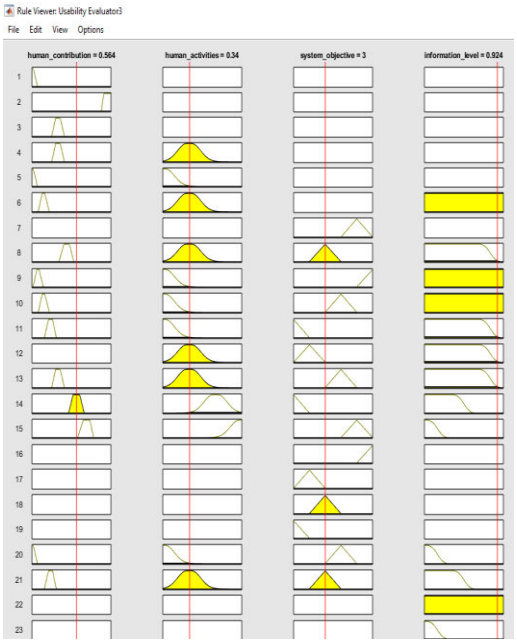


Fig. 11 Associated Inputs for an Interactive System in Fuzzy System

3.2.2 Set Related Evaluation Formula

We need to have MI, IR, and MV for the usability formula. In the example mentioned above, the obtained effect of usability factors (MI) is the output of the fuzzy system (Figure 12).

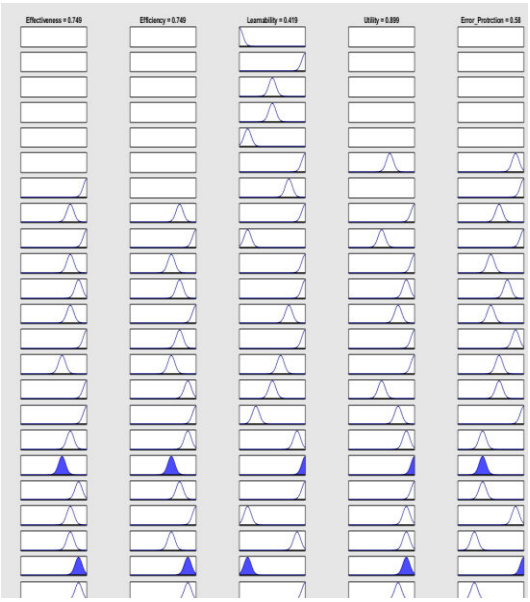


Fig. 12 Obtained Outputs in Sample Interactive System

When we have, a similar implemented, and usability tested interactive system, we select the proper RI from Table 1. In the aforementioned interactive system example, we assume not existing similar case so, the IR is 1 and the usability formula is (Equation 2),

$$Usability_{example} = 0.75\sqrt{MV_1} + 0.75\sqrt{MV_2} + 0.42\sqrt{MV_3} + 0.9\sqrt{MV_4} + 0.6\sqrt{MV_5} \quad (2)$$

3.2.3 Judgment Sampling

Expert judgment is the basis of evaluation usability values (MV). In the FIS system, a limited number of features can be considered. In this case, system designers must be experts in implementing and evaluating big data interactive systems. In our system, a judgmental sampling strategy is used to perform the evaluation. This method is a non-probability sampling method in which the researcher selects measurable features based on existing knowledge or professional judgment. In this sampling, the expertise of experts has priority instead of the number of experts [40]. In this study, we use the judgment sampling method to select experts. The experts jug the interactive system and provide the input of the fuzzy system and supply essential data for the multi-factor evaluation method.

3.2.4 Assess Metrics Based on The Fuzzy Calculating Distance

The implemented FIS is only used to determine the effect of usability metrics, but the value of usability metrics for evaluating an interactive system will obtain through an MFE method, which is an operational research approach. MFE typically deals with evaluating a set of alternatives based on multiple criteria and experts' judgment. Since MFE considers multiple factors for evaluation and it works based on decision makers' opinions, it can be used for the assessment software and systems when the empirical testing is complex. In this phase, we determine a multi-dimension scale corresponding to our criteria. We have five dimensions because of five usability metrics. The experts selected for judgment, indicate the performance of the entire interactive system in each metric with seven linguistic variables as Very bad (VB), Bad (B), Medium bad (MB), Medium (M), Medium good (MG), Good (G), Very good (VG). The fuzzy values for this scale based on triangular MF are (0, 0, 1), (0, 1, 3), (1, 3, 5), (3, 5, 7), (5, 7, 9), (7, 9, 10) and (9, 10, 10). We consider an ideal solution with a fuzzy value (10, 10, 10). The Metric Value (MV) is a distance between the ideal point and fuzzy value, it indicates by a subjective expert. The distance between these two triangular fuzzy numbers I = (I1, I2, I3) and M = (M1, M2, M3) is calculated according to fuzzy distance calculation presented in Equation 3 as:

$$MV = \sqrt{\frac{1}{3}[(I_1 - M_1)^2 + (I_2 - M_2)^2 + (I_3 - M_3)^2]} \qquad (3)$$

Where I = (10, 10, 10) and M is a fuzzy value that corresponds to linguistic variables, which are expressed by an expert for a metric.

3.3 Comparating Usability Between More Than One Interactive System

For comparing big data interactive systems with a different purpose, phase 2 is applicable as we run this phase separately for each interactive system and obtain usability scores. Then we look at the usability scores for comparing systems. However, when we want to compare big data interactive systems with the same purpose, human contribution, human activities, and information processing, we apply a pairwise comparison based MFE method for providing MVs. The usability of an interactive system with the same type calculates through Equation 4:

$$Usability_j = \sum_{i=1}^{5} \frac{\sqrt{MV_{ji} * MI_i^2}}{RI} \quad for\ j = 1\ to\ n \qquad (4)$$

For example, MV21 contains the value of the human contribution metric for interactive system 2. Also, when we have two big data interactive systems, then usability1 is the value of usability for interactive system one, and usability2 is the value of usability for interactive system 2.

3.3.1 Identify the Big data interactive systems Class

We have multiple interactive systems with the exact class of human contribution and activities, objectives, and information processing. Therefore, we will receive the same effect of usability metrics (MI) for these systems.

3.*3.2 Pairwise Comparating Systems on Related Criteria Based on the Fuzzy Variables*

The Pairwise comparison matrices are constructed to compare the interactive systems for each criterion (usability metric). The intensity of big data interactive systems in a metric corresponds using judgment's opinions through linguistic variables (Table 2).

TABLE 2 The linguistic Variable Scales and Their Related Fuzzy Numbers

Linguistic variables	Related Fuzzy Number
Very Strong (VS)	(7, 9, 10)
Fairly Strong (FS)	(5, 7, 9)
Strong (S)	(1, 3, 5)
Equal (E)	(1, 1, 1)
Weak (W)	(1, 1/ 3, 1/ 5)
Fairly Weak (FW)	(1/ 5, 1/ 7, 1/ 9)
Very Weak (VW)	(1/ 7, 1/ 9, 1/ 10)

The relative intensity of one system over another system for ranking in a usability metric is expressed using pairwise comparisons. These comparisons construct five pairwise comparison matrices corresponding to five criteria (usability metrics). Let $C = [\mathrm{C_i}]_\mathrm{n}$ $\mathrm{i} = 1, 2, \ldots, \mathrm{n}$ be the set of big data interactive systems. The result of the pairwise comparison is summarized in an evaluation matrix as follows (Equation 5):

$$CW = \begin{bmatrix} \mathrm{cw_{11}} & \ldots & \mathrm{cw_{1n}} \\ \vdots & \ddots & \vdots \\ \mathrm{cw_{n1}} & \ldots & \mathrm{cw_{nn}} \end{bmatrix} \quad (5)$$

Where $CW = [\mathrm{cw_{ij}}]_{\mathrm{n}\times\mathrm{n}}$ and $\mathrm{cw_{ij}}$ shows the intensity of the system $\mathrm{C_i}$ over system $\mathrm{C_j}$ through defuzzificating fuzzy values.

3.3.3 Obtaining Eigenvector

We produce the eigenvector from the pairwise comparison matrices to determine the ranking of big data interactive systems in each metric. We apply squaring, summarization, and normalization operations on pairwise comparison matrixes to obtain the eigenvector (Equations 5 & 6):

1. Squaring pairwise comparison matrix and construct S as $S = [\mathrm{s_{ij}}]_{\mathrm{n}\times\mathrm{n}}$.
2. Summarization row elements of matrix S and construct $\overrightarrow{CS} = [\mathrm{cs_i}]_\mathrm{n}$ where:

$$\mathrm{cs_i} = \sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{S_{ij}} \quad (5)$$

3. Normalization vector $\overrightarrow{CS}$ to reach eigenvector $\overrightarrow{CN} = [\mathrm{cn_i}]_\mathrm{n}$ where:

$$cn_k = \frac{\mathrm{cs_k}}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{cs_i}} \quad (6)$$

4. Repeat steps 1-3 and compare the unique vector in each iteration with the previous step to make the difference between the special vectors much smaller. The last special vector is the priority vector.

Previous mathematical studies have shown that special vector solutions are the best approach to obtain priority rankings from the pairwise comparison matrix [54]. Therefore, values of big data interactive systems in each metric will obtain from the eigenvector $\overrightarrow{CN}$. The appropriate vector is the priority that MV represents for big data interactive systems. Each pairwise comparison matrix corresponds to one criterion. The specific vector obtained for each matrix, including the rank of the systems in a criterion, is considered.

4. EXPERIMENTAL EXAMPLE

The research utilizes the experiment to evaluate the proposed model. The correlation between the results of the experimental test and the proposed model, represents the efficiency of the proposed model for evaluating big data interactive systems. Researchers apply different methodologies in ASRs (Automatic Speech Recognition) to address various types of disabilities [41]; [42]; [43]. The AMLMS algorithms are frequently applied in these systems to recognize the speech of disabled people who suffered from dysarthria [44]; [45]; & [46]. We implement a multimodel ASR interactive system in four ways of classification data in each mode that applies one of the AMLMs. We evaluate the usability in four modes of the multi-model ASR system through i) proposed model and ii) experimental study, then the results compare with statistical analysis (Figure 13). The multimodal ASR system applies for recognizing continuous speech at the sentence level. The effect of usability metrics for multimodel ASR is determined through a FIS (phase 1). Active learning methods can be divid-

ed into four categories: (1) single-view, single-learner (SVSL); (2) single-view, multi-learner (SVML); (3) multi-view, single-learner (MVSL); and (4) multi-view, multi-learner (MVML) [47]. Multimodel ASR is based on the AMLMs (SVSL, SVML, MVSL, and MVML) and acts like four separate big data interactive systems. Weighting the usability metrics for multimodel ASR is conducted through the third phase. Based on judgment sampling, three experts in HCI, ISPD, and AMLMs were recruited to compare four system modes in multiple metrics.

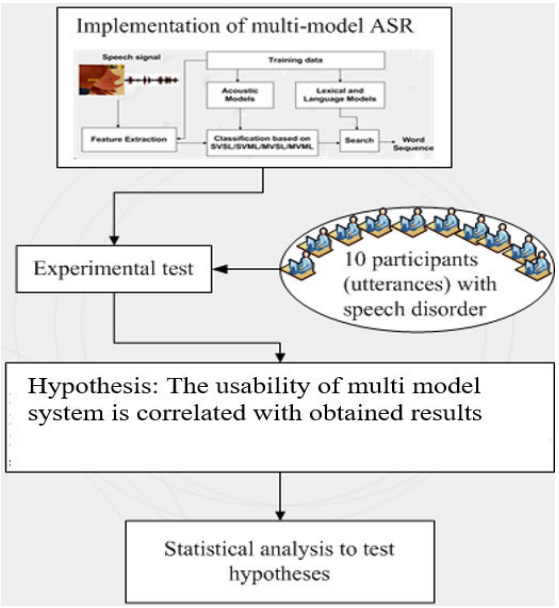


Fig. 13 Process of Experimental Testing

The experts' judgments are processed using Equations 2, 3, 4, 5, & 6 to obtain the usability score in four modes of the ASR system. In the experimental test, we recruited ten participants with speech disorders as utterances. The participants utter the sentences in different system modes then the usability metrics are measured based on system output. The correlation between the experimental results and the proposed model's results represents the efficiency of the proposed model for evaluating big data interactive systems. The hypotheses analysis is conducted to show the efficiency of the proposed model based on the obtained results in an experimental test of methods. The study population included academics working on HCIs, ISPDs, and machine learning techniques. The selection of these individuals was made by searching the search portal of academic researchers at http:/ / academic.research.microsoft.com/ and searching Google at http:/ / www.google.com. Finally, among the retrieved individuals, ten specialists appointed to collect data. After sending an electronic invitation to them, three experts responded positively and participated in the study. First, they briefly informed about the study background and objectives through the Skype video conference tool, and explanations gave about the multi-criteria evaluation method. Then, these experts were asked to determine the modes together according to five criteria and compare them based on the third step or study.

4.1 Multi-Model ASR System

This study has developed a multi-model ASR system in four modes; each mode applies one of the AMLMs to perform the data classification. This system as modern state of the art ASR system can act for pre-processing or feature extraction as well as acoustic, lexical, and language models. Some procedures have been developed for acoustic modeling, including Dynamic Time Warping (DTW), Hidden Markov Model (HMM), Vector Quantization, and Neural Networks [48]. HMM is a model based on the dominant recognition paradigm, in which speech changes are statistically modeled. In a multimodal ASR system, neural networks are commonly used to estimate word probabilities. In this system, the probabilities determined using HMM eventually become the most probable strings of the word. Figure 14 shows the general structure of the multi-model ASR system.

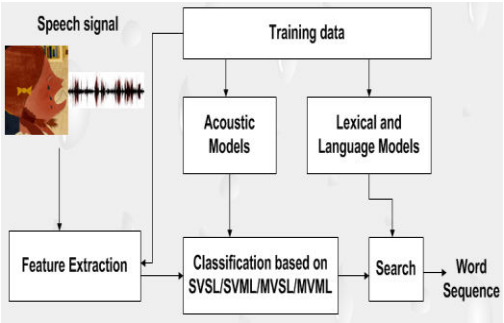


Fig. 14 Multi-Model ASR System

The ASR system detects several models of continuous speech at the first level (one-sentence phrases) for each participant. This system uses users' voice to detect users' speech disorders. These disorders include the production of incorrect speech sounds or improper sounds. Each participant spoke 215 sentences including a total of 1,079 words. Each sentence was spoken three times to minimize errors in mispronunciation in recording their speech. The audio is recorded in a suitable studio to reduce external noises. The sampling rate used to record was 16 kHz. All recorded voices were then labeled using a wave tagging tool to show the boundary of silence and pause between the words in each sentence. The MFCC technique was used to perform feature vector extraction operations using 25 ms frames and ten ms in the Hamming window.

4.2 Participants

We recruited ten participants with speech disorders. The participants have speech problems such as difficulties pronouncing sounds, or articulation disorders, and stuttering. Since there are two criteria (utility and Efficiency) subjective and must be rated based on participants' opinions, the participants can express their opinion in these two criteria are selected from previous studies. The demographic information of these participants is given in Table 3. The participants uttered 30 sentences containing 75 words (an average of 2.5 words per sentence) in different system modes. Then, the criteria are measured based on system output.

TABLE 3. Demographic Information of Participants

Participant	Gen.	Age	problem
1	F	24	pronouncing sounds
2	M	30	Stuttering
3	F	28	articulation disorders
4	F	27	Researcher pronouncing sounds
5	M	31	Stuttering
6	F	33	Researcher articulation disorders
7	F	35	Stuttering
8	M	25	pronouncing sounds
9	M	32	pronouncing sounds
10	M	38	articulation disorders

4.2.1 Metrics Measurements

System utility is measured subjectively (self-reported) and objectively (computer recorded measures). In some studies, using subjective measures is a weak form of measurement [49]. Using ASR show that some users' standard features in speeching are the repetition of some words, as well as the non-use of some words. However, using a specific system to conduct studies in this area, relies on the participants in the study. A multi-model ASR system depends on the number of times the system is used as an input intermediary and the tasks assigned to the system [50]. Learnability is a quantitative criterion calculated using the following formula (Equation 7):

$$\text{Learnability} = \left[\log_2 \sum_{i=1}^{n} \frac{(1 - \text{number of speech rejection})}{\text{Idealfeedback}_i}\right] \times 100 \quad (7)$$

Where n is the number of evaluation samples.

Error Protection is calculated through three elements: failure to hear or understand (E1); falsehoods produced in hearing or understanding (E2) and clarifications required to hear or understand (E3). These problems need to solve for both users, and the system (Equation 8).

$$\text{Error Handling} = \sqrt{\frac{\sum_{j=1}^{m}\sum_{i=1}^{n}\begin{bmatrix}((\text{IdealE1Output}_i - \text{ASRE1Output}_i)^2)_j + \\ ((\text{IdealE2Output}_i - \text{ASRE2Output}_i)^2)_j + \\ ((\text{IdealE3Output}_i - \text{ASRE3Output}_i)^2)_j\end{bmatrix}}{n \times m}} \quad (8)$$

In which m is the number of evaluation samples, and n is the vocabulary size.

The Efficiency indicated the ability of ASR to deal with various information attributes. We utilized a list of information attributes relevant for a broad spectrum of HCI applications that is presented by van and his colleagues [51]. It is highly subjective and commonly rated based on users' opinions. We then place each of the AMLMs for Audition Speech modality as "less," "neutral," or "more" appropriate to present the specific information attribute that is proposed by van Erp & Toet (2015)[51]. The amount of effectiveness in the ASR system usually indicates the accuracy of the tasks performed by the system, and this means the degree of accuracy of the system in detecting the user's speech. Accuracy is determined by checking the number of words caught. The percentage of these words is determined by the total number of words. Another alternative measure of the ASR system's response, especially when recognizing impaired speech, is the word error rate (WER) [50], which is formulated as follows (Equation 9):

$$\text{WER} = \frac{\text{Addition} + \text{Substitution} + \text{Omission}}{\text{Number of Words}} \text{X}100\% \quad (9)$$

Where:

- Phoneme addition is an extra sound (or sounds) added to the intended word
- Phoneme substitution is one phoneme substituted for another.
- Phoneme omission is a specific sound (or sounds) not produced.

Based on Saz et al. (2009) the Effectiveness is formulated as follows (Equation 10) [52]:

$$\text{Effectiveness} = 100\% - \text{WER} \quad (10)$$

4.2.2 Hypothesis Testing

The following hypotheses are defined to validate the outcomes obtained through the proposed model statistically: H1 – The usability of the multi-model system is correlated with obtained results
The list of variables used to test the hypothesis included:

1. Percentage of system performance (dependent, ratio)
2. Criteria (independent, nominal)
3. Method (independent, nominal)

In SPSS software, the Pearson correlation coefficient (CC) analyzes the relationship between rankings generated by the MFE method and experimental tests. This coefficient is a statistical tool to determine the type and extent of the relationship between one quantitative variable and another quantitative variable and shows the correlation between two variables [53]. Here, this method is used to determine the correlation between two variables. The correlation coefficient (r) shows how the data of a scatter are placed in a straight line.

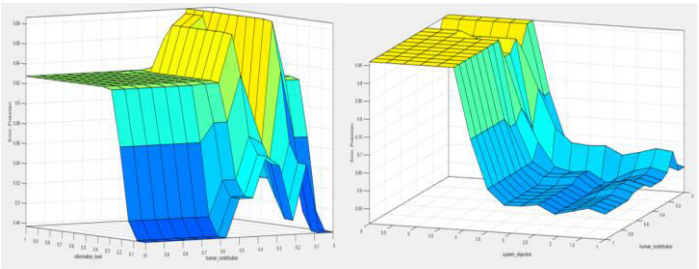
5. RESULTS AND DISCUSSION

The results of the study present in three sections. The first section is about the relation of interactive system classes and usability metrics, the second section is related to fuzzy MFE results, and the third

section is related to hypothesis MFE results.

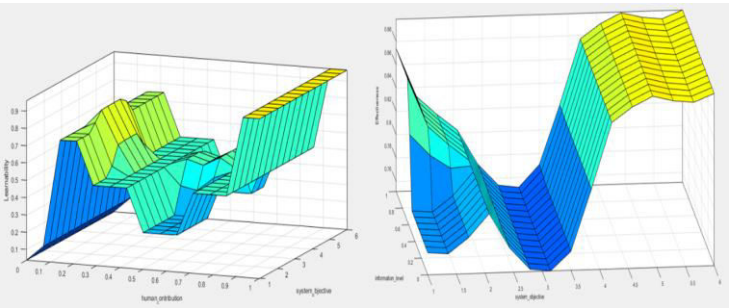
5.1 Relation of Interactive System Classes and Usability Metrics

In Figure 15, a part of produced output surfaces of the FIS according to set rules are demonstrated. Inferencing of fuzzy rules shows that at the low level of human contribution the error protection metric has higher importance than the high level of human contribution (Figure 15.a). Also, the system objective is associated with the effect of error protection where medical, and military have the highest importance of error protection (Figure 15.b). Human participation and the ability to learn are directly related to each other. Increasing human participation leads to increased learning ability. However, the system objective and learnability do not have direct connection (Figure 15.c). System objective has a direct relation with effectiveness; for example, medical has the highest effect of significance. The information processing level affects the effectiveness. The higher level of information processing leads to a higher effect on the effectiveness (Figure 15.d).



a. Relation of human contribution, information level, and error protection

b. Relation of system objective, human contribution, and error protection



c. Relation of human contribution, system objective, and learnability

d. Relation of system objective, information level, and effectiveness

Fig.15 A part of output surfaces

5.2 Fuzzy MFE Results

In the third phase, we used MFE methods for evaluating AMLMs. This evaluation is based on expert opinions rather than experiments. The experts have been selected from the academic board of the University of Malaya with experience, and knowledge in two scopes: i) AMLMs, and ii) ISPDs. Five evaluation criteria as “utility,” “Efficiency,” “Learnability,” “Effectiveness,” and “Error Protection” are determined in the first phase. The group experts were asked to compare the ASR modes with each other in the criteria (usability metrics). The aggregating their opinions is illustrated in Table 4.

We constructed the pairwise comparison matrix for each usability metric. We presented the results associated with utility metrics. Table 5 shows the comparison matrix associated to criterion “utility.”

TABLE 4. Pairwise Comparing ASR Modes

METHOD	Criteria	Comments	METHOD	Criteria	Comments
MVML vs MVSL	Utility	**MVML is FS in comparison with MVSL**	**MVSL vs SVML**	Utility	**MVSL is FS in comparison with SVML**
	Learnability	**MVML is FW in comparison with MVSL**		Learnability	**MVSL is FW in comparison with SVML**
	Error Protection	**MVML is FW in comparison with MVSL**		Error Protection	**MVSL is FW in comparison with SVML**
	Efficiency	**MVML is FW in comparison with MVSL**		Efficiency	**MVSL is FW in comparison with SVML**
	Effectiveness	**MVML is FW in comparison with MVSL**		Effectiveness	**MVSL is FS in comparison with SVML**
MVML vs SVML	Utility	**MVML is FS in comparison with SVML**	**MVSL vs SVSL**	Utility	**MVSL is E in comparison with SVSL**
	Learnability	**MVML is FS in comparison with SVML**		Learnability	**MVSL is E in comparison with SVSL**
	Error Protection	**MVML is FW in comparison with SVML**		Error Protection	**MVSL is E in comparison with SVSL**
	Efficiency	**MVML is FW in comparison with SVML**		Efficiency	**MVSL is FW in comparison with SVSL**
	Effectiveness	**MVML is FW in comparison with SVML**		Effectiveness	**MVSL is W in comparison with SVSL**
MVML vs SVSL	Utility	**MVML is VS in comparison with SVSL**	**SVML vs SVSL**	Utility	**SVML is FS in comparison with SVSL**
	Learnability	**MVML is VS in comparison with SVSL**		Learnability	**SVML is S in comparison with SVSL**
	Error Protection	**MVML is FW in comparison with SVSL**		Error Protection	**SVML is S in comparison with SVSL**
	Efficiency	**MVML is FW in comparison with SVSL**		Efficiency	**SVML is FS in comparison with SVSL**
	Effectiveness	**MVML is FW in comparison with SVSL**		Effectiveness	**SVML is FS in comparison with SVSL**

TABLE 5. Pairwise Comparison Matrix Related to the Utility Through Linguistic Variables

ALMs	**MVML**	**MVSL**	**SVML**	**SVSL**
MVML	E	FS	FS	VS
MVSL	-	E	FS	E
SVML	-	-	E	FS
SVSL	-	-	-	E

We replaced the linguistic variables with their corresponding fuzzy numbers determined in Table 2. Table 6 shows the fuzzified comparison matrix of utility.

TABLE 6. Fuzzy Pairwise Comparison Matrix Related to Utility

ALMs	**MVML**	**MVSL**	**SVML**	**SVSL**
MVM	(1,1,1)	(5,7,9)	(5,7,9)	(7,9,10)
MVSL	-	(1,1,1)	(5,7,9)	(1,1,1)
SVML	-	-	(1,1,1)	(5,7,9)
SVSL	-	-	-	(1,1,1)

Equation 2 has been applied for defuzzificating the comparison matrix of utility (Table 7).

TABLE 7. Defuzzified Pairwise Comparison Matrix Related to Utility

ALMs	**MVML**	**MVSL**	**SVML**	**SVSL**
MVML	1	7	7	8.75
MVSL	0.142857	1	7	1
SVML	0.142857	0.142857	1	7
SVSL	0.114286	1	0.142857	1

We obtained the eigenvector of the defuzzified pairwise comparison matrix related to utility. It is considered the MVs of ASR modes in utility criterion (Table 8).

TABLE 8. Usability Metric Values (MVs) in Utility

ASR Modes	**MV**
MVML	0.609701
MVSL	0.209054
SVML	0.11857
SVSL	0.0626755

We used the same procedure for obtaining the MVs in other criteria (Table 9).

TABLE 9. MVs of ASR Modes in all Criteria

Alm	Utility	Learnability	Error Protection	Efficiency	Effectiveness
MVML	0.60970	0.34056	0.04279	0.036215	0.02648
MVSL	0.20905	0.30748	0.16921	0.093077	0.44744
SVML	0.11857	0.29441	0.59382	0.632602	0.31333
SVSL	0.06267	0.05753	0.19417	0.238106	0.21273

MVML mode has the highest effect on utility and learnability. However, it has the lowest effect on error protection, efficiency, and effectiveness. SVML has the maximum effect on Error Protection and Efficiency. On the other hand, MVSL has the maximum effect on Effectiveness. The overall usability of ASR modes is calculated through Equations 2, 3, 4, 5, & 6 that as demonstrated in Table 10. The SVML mode has the maximum overall effect and, MVML has the third priority for use in ISPDs.

TABLE 10. Overall Effects of Methods

AML Method	**Effect**
SVML	1.952746
MVSL	1.226272
MVML	1.055758
SVSL	0.765226

5.3 Hypothesis Results

Through statistical analysis, we first proved that different ASR modes have different applications. The usability of this system is compared by the proposed model and experimental test. The overall results of the empirical tests and the average of users' answers are presented in Figure 16. The MVML mode has the maximum utility and learnability, the MVSL has the maximum effectiveness, SVML has the maximum error protection and efficiency.

The analysis of SPANOVA results shows that there is an interaction effect between ASR modes and usability score [$F(3, 36) = 76.926$, $p < .05$] (Table 11). The results of tests for subject effects indicate that there are significant differences between the modes in the overall usability. Based on Estimated Marginal Means (Table 12), the MVSL has better usability than other AMLMs.

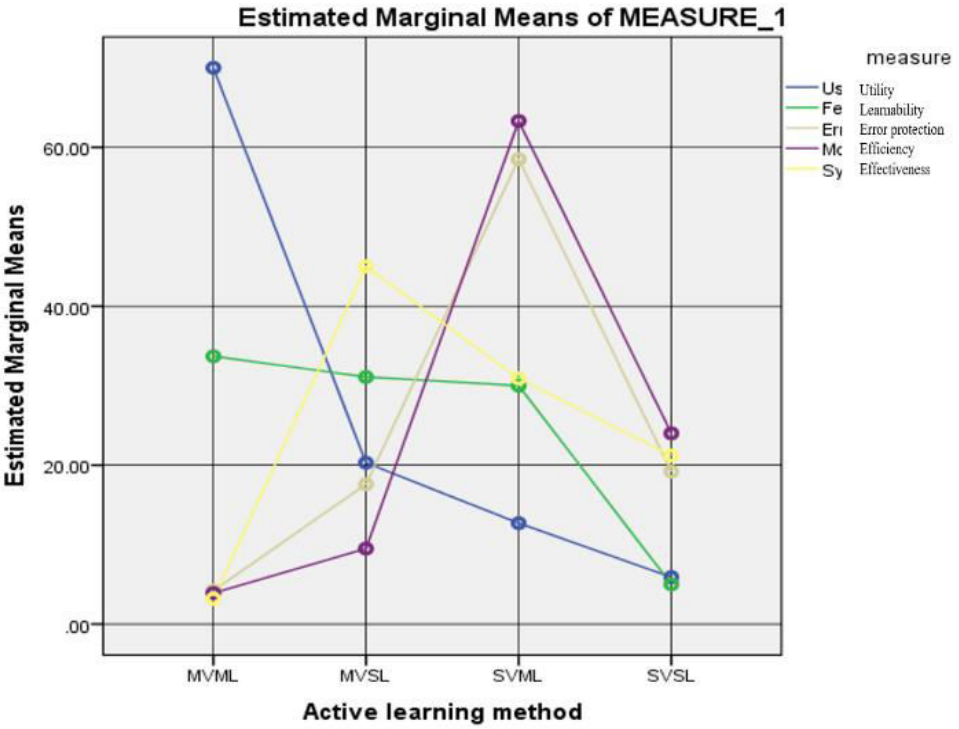


Fig.16 Overall Usability of ASR Modes

TABLE 11. Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	129642.320	1	129642.320	1992.624	.000
methods	15014.680	3	5004.893	76.926	.000
Error	2342.200	36	65.061		

TABLE 12. Usability Comparing ASR Modes

ASR Modes	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
MVML	23.000	1.141	20.687	25.313
MVSL	24.700	1.141	22.387	27.013
SVML	39.080	1.141	36.767	41.393
SVSL	15.060	1.141	12.747	17.373

The results of Post Hoc Tests show that there is not a significant difference between MVSL and MVML in usability (Table 13). Also, the classification results of methods show that MVML and MVSL are in the same group in terms of their effectiveness (Table 14).

The CC analysis is conducted to find the relation between the ranking of modes produced by the proposed model and the empirical system test. The results show that there is a strong and linear relation between them. N is the number of criteria considered in rankings (Table 15).

TABLE 13. Multiple Comparisons

Tukey HSD				
(I) ASR mode	(J) ASR mode	Mean Difference (I-J)	Std. Error	Sig.
MVML	MVSL	-1.7000	1.61321	.719
	SVML	-16.0800*	1.61321	.000
	SVSL	7.9400*	1.61321	.000
MVSL	MVML	1.7000	1.61321	.719
	SVML	-14.3800*	1.61321	.000
	SVSL	9.6400*	1.61321	.000
SVML	MVML	16.0800*	1.61321	.000
	MVSL	14.3800*	1.61321	.000
	SVSL	24.0200*	1.61321	.000
SVSL	MVML	-7.9400*	1.61321	.000
	MVSL	-9.6400*	1.61321	.000
	SVML	-24.0200*	1.61321	.000

TABLE 14. Homogeneous Subsets

ASR mode	N	Subset		
		1	2	3
SVSL	10	15.0600		
MVML	10		23.0000	
MVSL	10		24.7000	
SVML	10			39.0800
Sig.		1.000	.719	1.000

TABLE 15. Correlation matrix of MFE method and Empirical system test

		MFE method	Empirical test
MFE method	Pearson Correlation	1	.928**
	Sig. (2-tailed)		.000
	N	5	5
Empirical test	Pearson Correlation	.928**	1
	Sig. (2-tailed)	.000	
	N	5	5
**. Correlation is significant at the 0.01 level (2-tailed).			

6. CONCLUSION

The usability evaluation by the big data interactive systems is a decision-making issue and it has a strong influence on the overall improvement of big data interactive systems. In the formative evaluation of an interactive system or situations that conducting empirical tests are costly, the developers need to predict the usability without conducting empirical tests. The usability evaluation of big data interactive systems considers in terms of qualitative and quantitative criteria. Moreover, there is no constant situation for evaluating big data interactive systems, the importance of usability metrics will change based on the interactive system. Therefore, an efficient and dynamic evaluation method is necessary to improve the evaluation process. In this study, we proposed an integrated model with three phases of evaluation. The fuzzy method is integrated with MFE methods to increase the accuracy of evaluation. The first phase was the preprocessing of evaluation. In this phase, we determined the usability factors based on the most associated standards and literature review. Four classifications of big data interactive systems are proposed based on a survey of 182 associated articles. We implemented a FIS with 50 fuzzies if-then rules to produce the best metric effect for any interactive system. Two fuzzy MFE methods are proposed for i)

evaluating an interactive system based on fuzzy distance calculation to ideal solution and ii) comparing multiple big data interactive systems based on the pairwise comparison, along with two formulations of overall usability correspondingly. To the best of our knowledge, an expert-based usability evaluation fuzzy system is a novel system with a dynamic aspect. The fuzzification scale of linguistic variables design based on the experts' opinions. The proposed model is applied to assess the usability of an implemented multimodel ASR for people with disabilities. The results show that the proposed model has an accurate prediction of usability scores for four modes of the ASR system. The MVML mode had the highest effect on utility and learnability. However, it had the lowest value for error protection, efficiency, and effectiveness. SVML mode had the maximum value in Error Protection and Efficiency. The results of the proposed model have been examined through experimental and hypothesis tests. The system is tested for participants with speech disorders. The criteria are measured separately in four modes based on the system's performance. The statistical results show the importance of AMLMs. The hypotheses analyses indicated the high correlation between proposed model results and experimental results. The output surfaces of the fuzzy systems allowed us to determine the relationship between interactive system types and usability metrics. It is concluded that applying a dataset in a neuro-FIS and training system cause to produce more than a hundred effective rules. The findings indicate that the proposed model can apply for interactive system evaluation, informative evaluation, and conducting complex empirical tests. Future studies may improve the FIS with the integration of artificial neural networks.

7. Declarations

- Ethics approval and consent to participate: This article does not contain any studies with human or animal participants performed by any of the authors.
- Consent for publication: The publisher has the author's permission to publish the work.
- Availability of data and materials: Available if needed
- Conflict of Interest: No Conflict of Interest
- Funding: Not applicable
- Authors contribution: Two authors contributed to the manuscript equally
- Acknowledgments: “This research has been supported by the ’Potential, Forms, Economic Value and Social Benefits of Innovation in the Digital Society’ program of the Corvinus University of Budapest financed by the 2019 Domain Excellence Fund of the National Research, Development and Innovation Office”.

8. References

[1] Otero, C. (2012). Software engineering design: theory and practice: CRC Press. ISBN 9781439851685.

[2] An, Q., Wang, Z., Emrouznejad, A., Zhu, Q., & Chen, X. (2019). Efficiency evaluation of parallel interdependent processes systems: An application to Chinese 985 Project universities. *International Journal of Production Research*, *57*(17), 5387–5399. https://doi.org/10.1080/00207543.2018.1521531

[3] Benson, M. S., & Powell, M. B. (2015). Evaluation of a comprehensive interactive training system for investigative interviewers of children. *Psychology, Public Policy, and Law*, *21*(3), 309–322. https://doi.org/10.1037/law0000052

[4] Pal, D., & Vanijja, V. (2020). Perceived usability evaluation of Microsoft Teams as an online learning platform during COVID-19 using system usability scale and technology acceptance model in India. *Children and Youth Services Review*, *119*, 105535. https://doi.org/10.1016/j.childyouth.2020.105535

[5] Singla, B. S., & Aggarwal, H. (2020). A Set of Usability Heuristics and Design Recommendations for Higher Education Institutions' Websites. International Journal of Information System Modeling and Design (IJISMD), 11(1), 58-73. https://doi.org/10.4018/ijismd.2020010104

[6] Azeem, M., Ahmed, M., Sajid, A., Iqbal, T., Farooq, A., & Valsalan, P. (2019). An Interactive Code Generator and Profiler System. International Journal of Computer Science and Network Security, 19(1), 223-231. http://paper.ijcsns.org/07_book/201901/20190127.pdf

[7] Durango, I., Carrascosa, A., Gallud, J. A., & Penichet, V. M. R. (2018). Interactive fruit panel (IFP): a tangible serious game for children with special needs to learn an alternative communication system. Universal Access in the Information Society, 17(1), 51-65. https://doi.org/10.1007/s10209-016-0517-5

[8] Ganjali, R., Taherzadeh, Z., Sabbagh, M. G., Nazemiyan, F., Mamdouhi, F., Tabesh, H., . . . Eslami, S. (2019). Effect of an interactive voice response system on self-management in kidney transplant recipients: Protocol for a randomized controlled trial. Medicine, 98(6). https://doi.org/ARTN e1429110.1097/MD.0000000000014291

[9] Khowaja, K., Salim, S. S., & Asemi, A. (2015). Heuristics to evaluate interactive systems for children with autism spectrum disorder (ASD). PloS one, 10(7), e0132187.

[10] Koutsabasis, P. (2017). Empirical Evaluations of Interactive Systems in Cultural Heritage: A Review. *International Journal of Computational Methods in Heritage Science*, *1*(1), 100–122. https://doi.org/10.4018/IJCMHS.2017010107

[11] Ofli, F., Kurillo, G., Obdržálek, Š., Bajcsy, R., Jimison, H. B., & Pavel, M. (2016). Design and evaluation of an interactive exercise coaching system for older adults: lessons learned. IEEE journal of biomedical and health informatics, 20(1), 201-212.

[12] Santana, P. C., & Anido, L. E. (2016). Heuristic evaluation of an interactive television system to facilitate elders home care. IEEE Latin America Transactions, 14(7), 3455-3460.

[13] Hartson, H. R., Andre, T. S., & Williges, R. C. (2003). Criteria for evaluating usability evaluation methods. International Journal of Human-Computer Interaction, 15(1), 145-181.

[14] MacDonald, C. M., & Atwood, M. E. (2013). Changing perspectives on evaluation in HCI: past, present, and future. Paper presented at the CHI'13 extended abstracts on human factors in computing systems.

[15] Nielsen, J. (1994). Usability Inspection Methods. *Celebrating Interdependence CH94*, 413–414. https://rauterberg.employee.id.tue.nl/lecturenotes/0H420/Nielsen[1994].pdf

[16] Lee, Y., & Lee, J. (2019). A checklist for assessing blind users' usability of educational smartphone applications. Universal Access in the Information Society, 18(2), 343-360. https://doi.org/10.1007/s10209-017-0585-1

[17] Veral, R., & Macías, J. (2019). Supporting user-perceived usability benchmarking through a developed quantitative metric. Int. J. Hum. Comput. Stud. https://doi.org/10.1016/j.ijhcs.2018.09.012

[18] Othman, M. K., Sulaiman, M. N. S., & Aman, S. (2018). Heuristic Evaluation: Comparing Generic and Specific Usability Heuristics for Identification of Usability Problems in a Living Museum Mobile Guide App. Advances in Human-Computer Interaction. https://doi.org/10.1155/2018/1518682

[19] Yanez-Gomez, R., Font, J. L., Cascado-Caballero, D., & Sevillano, J. L. (2019). Heuristic usability evaluation on games: a modular approach. Multimedia Tools and Applications, 78(4), 4937-4964. https://doi.org/10.1007/s11042-018-6593-1

[20] Memedi, M., Tshering, G., Fogelberg, M., Jusufi, I., Kolkowska, E., & Klein, G. (2018). An Interface for IoT: Feeding Back Health-Related Data to Parkinson's Disease Patients. Journal of Sensor and Actuator Networks, 7(1). https://doi.org/10.3390/jsan7010014

[21] Rahmat, H., Zulzalil, H., Abd Ghani, A. A., & Kamaruddin, A. (2018). A Comprehensive Usability Model for Evaluating Smartphone Apps. Advanced Science Letters, 24(3), 1633-1637. https://doi.org/10.1166/asl.2018.11125

[22] Imel, Z. E., Pace, B. T., Soma, C. S., Tanana, M., Hirsch, T., Gibson, J., . . . Atkins, D. C. (2019). Design Feasibility of an Automated, Machine-Learning Based Feedback System for Motivational Interviewing. Psychotherapy, 56(2), 318-328. https://doi.org/10.1037/pst0000221

[23] Zadeh, L. A. (1965). "Fuzzy sets." Information and Control 8(3): 338-353.

[24] Zadeh, L. A. (2015). Fuzzy logic-a personal perspective. Fuzzy Sets and Systems, 281, 4-20. https://doi.org/10.1016/j.fss.2015.05.009

[25] Alenezi, M., Agrawal, A., Kumar, R., & Khan, R. A. (2020). Evaluating performance of Web application security through a fuzzy based hybrid multi-criteria decision-making approach: Design tactics perspective. IEEE Access, 8, 25543-25556.

[26] Chang, C., & Almaghalsah, H. (2020). Usability evaluation of e-government websites: A case study from Taiwan. International Journal of Data and Network Science, 4(2), 127-138.

[27] Dawood, K. A., Sharif, K. Y., Ghani, A. A., Zulzalil, H., Zaidan, A., & Zaidan, B. (2021). Towards a unified criteria model for usability evaluation in the context of open-source software based on a fuzzy Delphi method. Information and Software Technology, 130, 106453.

[28] Ramanayaka, K. H., Chen, X. Q., & Shi, B. (2019). Application of extent analysis FAHP to determine the relative weights of evaluation indices for library website usability acceptance model. Iet Software, 13(1), 86-95. https://doi.org/10.1049/iet-sen.2018.5185

[29] Chang, E., & Dillon, T. S. (2006). A usability-evaluation metric based on a soft-computing approach. Ieee Transactions on Systems Man and Cybernetics Part a-Systems and Humans, 36(2), 356-372. https://doi.org/10.1109/tsmca.2005.851349

[30] Chang, E. J., Dillon, T. S., & Cook, D. (1997). An intelligent system-based usability evaluation metric. https://doi.org/10.1109/IIS.1997.645226.

[31] Kumar, A., Tadayoni, R., & Sorensen, L. T. (2015). Metric Based Efficiency Analysis Of Educational ERP System Usability-Using Fuzzy Model.

[32] Huddy, J. R., Ni, M., Misra, S., Mavroveli, S., Barlow, J., & Hanna, G. B. (2019). Development of the Point-of-Care Key Evidence Tool (POCKET): a checklist for multi-dimensional evidence generation in point-of-care tests. Clinical Chemistry and Laboratory Medicine, 57(6), 845-855. https://doi.org/10.1515/cclm-2018-1089

[33] Sheridan, T. B., & Verplank, W. L. (1978). Human and computer control of undersea teleoperators. https://apps.dtic.mil/sti/pdfs/ADA057655.pdf

[34] Sharp, H., Preece, J., & Rogers, Y. (2019). Interaction Design: Beyond Human-Computer Interaction (J. W. S. Inc Ed. 5 ed.). ISBN: 978-1-119-54730-3.

[35] Giuzio, A., Mecca, G., Quintarelli, E., Roveri, M., Santoro, D., & Tanca, L. (2019). INDIANA: An interactive system for assisting database exploration. Information Systems, 83, 40-56. https://doi.org/10.1016/j.is.2019.01.003

[36] Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and groups of human interaction with automation. IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans, 30(3), 286-297.

[37] Bevan, N., Carter, J., & Harker, S. (2015). ISO 9241-11 revised: What have we learnt about usability since 1998? Paper presented at the International Conference on Human-Computer Interaction.

[38] Mandal, S. N., Choudhury, J. P., & Chaudhuri, S. B. (2012). In search of suitable fuzzy MF in prediction of time series data. International Journal of Computer Science Issues, 9(3), 293-302. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.402.1043&rep=rep1&type=pdf

[39] Noble, E. E., & Sanchez, P. P. (1993). A note on the information content of a consistent pairwise comparison judgment matrix of an AHP decision maker. Theory and Decision, 34(2), 99-108.

[40] Taherdoost, H. (2016). Sampling methods in research methodology; How to choose a sampling technique for research.

[41] Hornof, A., Whitman, H., Sutherland, M., Gerendasy, S., McGrenere, J., & Acm. (2017). Designing for the "Universe of One": Personalized Interactive Media Systems for People with the Severe Cognitive Impairment Associated with Rett Syndrome.

[42] Scott, R. A., Callisaya, M. L., Duque, G., Ebeling, P. R., & Scott, D. (2018). Assistive technologies to overcome sarcopenia in ageing. Maturitas, 112, 78-84. https://doi.org/10.1016/j.maturitas.2018.04.003

[43] Wagar, M. M., Aslam, M., & Farhan, M. (2019). An Intelligent and Interactive Interface to Support Symmetrical Collaborative Educational Writing among Visually Impaired and Sighted Users. Symmetry-Basel, 11(2). https://doi.org/10.3390/sym11020238

[44] Shahamiri, S. R. and S. Binti Salim (2014a). "A Multi-Views Multi-Learners Approach Towards Dysarthric Speech Recognition Using Multi-Nets Artificial Neural Networks." IEEE Transactions on Neural Systems and Rehabilitation Engineering 22 (5): 1053 - 1063

[45] Shahamiri, S. R. and S. S. Binti Salim (2014b). "Artificial neural networks as speech recognisers for dysarthric speech: Identifying the best-performing set of MFCC parameters and studying a speaker-independent approach." Advanced Engineering Informatics 28(1): 102-110.

[46] Shahamiri, S. R. and S. S. Binti Salim (2014c). "Real-time frequency-based noise-robust Automatic Speech Recognition using Multi-Nets Artificial Neural Networks: A multi-views multi-learners approach." Neurocomputing 129: 199-207.

[47] Jamshidpour, N., Safari, A., & Homayouni, S. (2020). A GA-Based Multi-View, Multi-Learner Active Learning Framework for Hyperspectral Image Classification. Remote Sensing, 12(2), 297. https://doi.org/10.3390/rs12020297

[48] Chugh, A., Jain, C., Singh, P., & Rana, P. (2015). Learning approach for offline signature verification using vector quantization technique. In Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1 (pp. 337-344). Springer International Publishing.

[49] Szajna, B., & Scamell, R. W. (1993). The effects of information system user expectations on their performance and perceptions. Mis Quarterly, 493-516. https://doi.org/10.2307/249589.

[50] Koester, H. H. (2004). Usage, performance, and satisfaction outcomes for experienced users of automatic speech recognition. *Journal of Rehabilitation Research and Development*, *41*(5), 739–754. https://doi.org/10.1682/jrrd.2003.07.0106

[51] van Erp, J. B. F., & Toet, A. (2015). Social Touch in Human–Computer Interaction. *Frontiers in Digital Humanities*, *0*. https://doi.org/10.3389/fdigh.2015.00002

[52] Saz, O., Yin, S.C., Lleida, E., Rose, R., Vaquero, C., & Rodrı´guez, W.R. (2009). Tools and technologies for computer-aided speech and language therapy. Speech Communication, 51, 948–967. https://doi.org/10.1016/ j.specom.2009.04.006

[53] Croxton, F. E., & Cowden, D. J. (1939). Applied general statistics.

[54] Kahraman, C., Ruan, D., & Doğan, I. (2003). Fuzzy group decision-making for facility location selection. Information Sciences, 157, 135-153.

[55] Ponce, P., Balderas, D., Peffer, T., & Molina, A. (2018). Deep learning for automatic usability evaluations based on images: A case study of the usability heuristics of thermostats. Energy and Buildings, 163, 111-120. https://doi.org/10.1016/j.enbuild.2017.12.043