

Article

A Weighted Bonferroni-OWA Operator Based Cumulative Belief Degree Approach to Personnel Selection Based on Automated Video Interview Assessment Data

Umüt Asan  and Ayberk Soyer 

Department of Industrial Engineering, Istanbul Technical University, Macka, Istanbul 34367, Turkey; ayberk@itu.edu.tr

* Correspondence: asanu@itu.edu.tr

Abstract: Asynchronous Video Interviewing (AVI) is considered one of the most recent and promising innovations in the recruitment process. Using AVI in combination with AI-based technologies enables recruiters/employers to automate many of the tasks that are typically required for screening, assessing, and selecting candidates. In fact, the automated assessment and selection process is a complex and uncertain problem involving highly subjective, multiple interrelated criteria. In order to address these issues, an effective and practical approach is proposed that is able to transform, weight, combine, and rank automated AVI assessments obtained through AI technologies and machine learning. The suggested approach combines Cumulative Belief Structures with the Weighted Bonferroni-OWA operator, which allows (i) aggregating assessment scores obtained in different forms and scales; (ii) incorporating interrelationships between criteria into the analysis (iii) considering accuracies of the learning algorithms as weights of criteria; and (iv) weighting criteria objectively. The proposed approach ensures a completely data-driven and efficient approach to the personnel selection process. To justify the effectiveness and applicability of the suggested approach, an example case is presented in which the new approach is compared to classical MCDM techniques.

Keywords: asynchronous video interviewing; personnel selection; multi-criteria decision making; cumulative belief structures; Bonferroni mean; ordered weighted averaging operator; machine learning; automated assessment

MSC: 03E72; 68T37; 90B50; 90B70

Citation: Asan, U.; Soyer, A. A Weighted Bonferroni-OWA Operator Based Cumulative Belief Degree Approach to Personnel Selection Based on Automated Video Interview Assessment Data. *Mathematics* **2022**, *10*, 1582. <https://doi.org/10.3390/math10091582>

Academic Editor: Antonio Jiménez-Martín

Received: 11 April 2022

Accepted: 5 May 2022

Published: 7 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recent technological advances offer the opportunity for the human resources (HR) function to redesign its processes. In particular, the recruitment and selection process has the potential for significant improvement [1]. The traditional recruitment and selection process that is still used in the HR field has been proven to have certain drawbacks such as (i) inaccessibility of candidates located in different geographic regions; (ii) inability to re-evaluate non-verbal cues; (iii) high travel expenses; (iv) difficulty in scheduling interview times; (v) inability to archive interviews; (vi) making recruitment decisions with a limited number of evaluators; and (vii) long recruitment times. Therefore, technology-based intelligent approaches are needed to provide flexibility to both candidates and organizations, and to ensure that vacancies are filled quickly and effectively at low cost [2,3].

More recently though, how recruiters work and job seekers search for jobs has begun to change. Globalization and the internet made physical distance much less relevant and allowed employers to carry out much of their recruitment process online [1]. Job boards/job sites, company career sites, and social networking websites are common platforms used for online recruitment [4]. However, these developments raise new challenges to be addressed. In particular, how to deal with much larger numbers of applications in terms of screening,

interviewing, and selection is a critical issue. Reducing the applicants pool, holding interviews, and selecting the most suitable candidate(s) all require a great deal of time and may exceed HR's working capacity [1,5].

Asynchronous Video Interviewing (AVI), which aims to address these challenges, is considered one of the most recent and promising innovations in the recruitment process [6–8]. Additionally, known as a digital or one-way interview in the literature, AVI is a type of interview in which candidates record (and upload) their answers to a predetermined set of interview questions without meeting or speaking with a human interviewer [4,9]. Since an interview video allows multiple playbacks, conveys much more information than text, and assesses candidates in a standardized way, AVI provides an effective solution for selecting promising candidates [10]. In addition, AVI allows for the provision of equal opportunity to candidates (especially to long-distance candidates), reducing cost per hire, shortening time to fill open vacancies, observing candidates' body language, preventing inherent biases in the interview process, and reducing the potential negative impact of memory [11,12].

Assessments of the recordings of the candidates rely mostly on the intuition of interviewers. However, using AVI in combination with other techniques and technologies such as automatic speech recognition (ASR), natural language processing (NLP), and machine learning (ML) offers further improvements in the assessment process. The latter enables recruiters/employers to automate many of the tasks that are typically required for personnel selection, such as screening and assessing candidates [1,13]. Another important benefit to employers is the increased capacity for dealing with a large number of candidates [7]. The automated assessment option of AVI platforms is more likely to be utilized in the early stages of the hiring process to screen for higher-quality candidates to be invited for an initial in-person interview [4,7]. Alternatively, in the traditional process, preliminary screening of eligibility is performed solely based on resumes and/or phone conversations, which provide only limited insight.

Intelligent hiring and AI-enabled AVI platforms such as HireVue [14], Modern Hire [15] (formerly Montage), and Talview [16] are gaining increased attention. Large companies around the world such as Unilever, IBM, Mercedes Benz, Walmart, L'Oréal, and Hilton have already adopted an AI-enabled AVI and assessment platform [1,9,17]. Unfortunately, in comparison to the growing use of AVI and automated assessment tools in practice, little research has been conducted on this topic [4]. Studies in the literature can be roughly categorized into two main research directions: (i) empirical studies concerning perceptions, fairness, flexibility, and use of AVI ([1,11,18], among others); and (ii) methodological studies concerning technology- and assessment-related issues. Because of their relevance to this study, only the second research direction will be discussed in this study.

Technology and assessment related studies suggest using statistical and AI-based approaches to learn and predict patterns in candidate video responses. These studies attempt to assess several different criteria such as job-related skills and personality traits based on verbal and non-verbal (e.g., gestural, facial) indicators of the candidates during the interviews. Because of the uncertain multi-criteria multimodal nature of the problem, many of the proposed technology-related approaches, especially all-in-one approaches, suffer from highly complex modeling and lack of large and adequate datasets (see [19–21]). For example, assessing five job-related skills and eight different personality traits of 100 candidates based on both verbal and non-verbal contents (tone of voice, facial expressions, and body gestures) requires modeling “verbal response—non-verbal behavior—score” patterns for several different skills, traits, and their respective questions (e.g., five questions for each skill). As in the given example, dealing with large numbers of candidates, involving a wide range of related skills, behavioral cues, and highly subjective responses, makes the automated assessment and prediction process highly complex ([11,12,21]). Considering all skills, behavioral cues, and questions in one single model or separately is a critical decision in terms of complexity, efficiency, accuracy, and data availability. The latter option is relatively simple, with the chance of more accurate results for even smaller datasets. This, in turn, raises the question of how to transform, weight, combine, and rank the

results of separate models (for different questions, skills, and/or modalities), which is one of the main concerns of this paper. Moreover, all-in-one models, especially those using sophisticated learning algorithms such as deep learning, are not able to provide candidates with explicit reasons for their results. Although explainable AI is on the way to solve this problem [18], explainability and accountability of automated selection algorithms will remain an important concern in the near future.

Another methodologically oriented research direction involves studies that treat the personnel assessment and selection process as an MCDM problem. These decision-making-related studies aim to model and analyze the problem by obtaining, weighting, and aggregating interviewer assessments based on multiple criteria. None of these studies report modeling of automated assessments of asynchronous video interviews. Instead, models developed in this category typically use subjective assessments obtained from classical face-to-face interviews and/or tests (e.g., personality, psychometric), and therefore may involve human bias. Relying on human raters decreases the efficiency of MCDM techniques that are based on comparisons, especially in case of large numbers of candidates. On the other hand, decision-making-related studies provide a number of advantages, such as (i) offering a large range of techniques for weighting and aggregating assessments; (ii) dealing with uncertainty and subjectivity involved in the assessment process; and (iii) considering both qualitative and quantitative data.

Although personnel selection is not a new topic in the MCDM literature, using assessments obtained from learning-based (data-driven) methods to make complex decisions is rather a new research direction. In other words, integrating learning-based techniques with MCDM techniques is gaining more interest [22]. To the best knowledge of the authors, there is no study reporting the use of automated AVI assessment data in MCDM models. AVI assessments as mentioned above involve multiple interrelated (sub-)criteria that can take various forms. In order to handle the complexity and uncertainty inherent in the assessment process, a flexible approach that is able to transform various forms of assessments into a common scale and aggregate them is needed.

To address these needs, this paper suggests an extended cumulative belief degree (CBD) approach to personnel selection based on automated AVI assessment scores. CBD as a flexible aggregation method [23] is proposed to transform these scores, which may be obtained in different forms from separate predictive models, into (cumulative) belief degrees. To aggregate cumulative belief degrees the Weighted Bonferroni-Ordered Weighted Averaging (wBON-OWA) operator is proposed. Using this objective weighting and averaging operator ensures a completely data-driven and computationally efficient approach to the whole selection process. Since CBD yields an aggregated score as a distribution over a predefined linguistic term set, it provides further insight into the suitability of a candidate. The proposed approach also incorporates the accuracy values of the learning algorithms as weights into the aggregation function. Using the generalized Bonferroni mean to capture the expressed interrelationship between the (sub-)criteria [24] is another significant contribution of this paper.

Consequently, the methodological and practical contributions of the suggested approach can be summarized as follows:

- The proposed approach is a prime example for the integration of learning-based techniques with MCDM techniques. This is the first study that transforms and aggregates automated AVI assessments in a multi-criteria environment for personnel selection.
- To the best knowledge of the authors, this is the first study that combines wBON-OWA with the CBD approach to consider (i) accuracies of the learning algorithms as weights of criteria; (ii) interrelationships between sub-criteria (i.e., interview questions); and (iii) objective weights of (sub-)criteria where higher (lower) scores are given more importance.
- Unlike the current MCDM approaches in the literature, a completely data-driven assessment and selection process without the need for any expert intervention is suggested.

- Depending on the ML algorithm used, the score obtained for each interview question can be in different forms and scales (e.g., probability, distance, test score, linguistic term). The proposed approach does not require any particular scaling properties of the data.
- Using a fuzzy linguistic term set to transform the scores into a common scale also allows consideration of the uncertainty inherent in the assessment process.
- Finally, the proposed approach effectively and efficiently copes with large numbers of (sub-)criteria and candidates.

The rest of the paper is structured as follows. In the next section, a review of related work is provided. In order to provide theoretical background for the proposed approach, the cumulative belief degree approach and the aggregation operators used are summarized in Section 3. The details of the proposed wBON-OWA-based cumulative belief degree approach are then presented in Section 4. In Section 5, an example case to illustrate the applicability and effectiveness of the proposed approach is presented. Finally, conclusions and future research directions are provided in the last section.

2. Related Works

Assessment and selection of candidates is not a new research topic and has been widely studied over the years. It is a complex and uncertain decision problem involving subjective interrelated multiple criteria. Here, two lines of research that are directly related to this study will be reviewed.

The first research direction involves studies, most of which aim to automate the whole assessment and selection process using AI technologies such as ASR, NLP, computer vision, and ML. Several automated assessment and prediction frameworks have been suggested based on a variety of online and offline methods available for interviewing candidates. Only recent studies based on automated assessment of asynchronous video interviews will be discussed here.

In AVI, candidates and interviewers do not meet. Interviewers upload predetermined interview questions for candidates to view and respond to at any time convenient to them within a predefined time interval. The candidates record themselves answering the given questions and submit their videos for assessment. AVI provides the opportunity to assess several different criteria, such as job-related skills and personality traits based on verbal and non-verbal indicators of the candidates, to learn and predict patterns in video responses. While verbal content involves what is being said and organization of ideas, non-verbal cues involve features such as tone of voice, facial expressions, and body gestures that help to recognize feelings, attitudes, and personality traits [8,20,25].

Methods proposed in the literature for automated assessment and prediction commonly involve the following stages (cf. [20]): (i) audio–visual recording of interviews; (ii) extraction of verbal and/or non-verbal features (i.e., unimodal or multimodal); and (iii) learning patterns to predict the job suitability of candidates either as a class through classification, or a continuous value through regression. Studies in the literature differ according to their contribution to one or more of these stages. For example, L. Chen, Feng, Martin-Raugh et al. [26] suggest an automated interview scoring system for monologue video interviews. The overall interview performance is predicted using ML algorithms based on a set of verbal (speech and lexical) and non-verbal (visual) features. Various feature-extraction tools and regression-based algorithms are used to model and predict the overall score. In another study by L. Chen, Feng, Leong et al. [13], an automatic video rating approach based on a new doc2vec feature-extraction method is proposed. In this approach, multimodal signals are clustered using K-means to visual words and documents consisting of these visual words are represented as feature vectors using doc2vec. The feature vectors are then used as input for ML algorithms. Similarly, Rasipuram et al. [27] develop an automated system to model and predict communication skills of candidates in AVI by extracting verbal and non-verbal behavioral cues and using regression and classification-based ML algorithms. A more recent study by L. Chen et al. [28] presents an

automated scoring approach to effectively converting and joining multimodal behaviors based on a large-sized corpus of 1891 monologue videos collected through Amazon Mechanical Turk workers. After converting candidates' audio responses to texts using ASR, Bag of Words feature representation and text-based classifiers are used to predict the overall performance and personality scores. In another study, Rao et al. [8] suggest a model using automatically extracted multimodal features (such as lexical, audio, and visual features) and classical ML methods to assess communication skills of candidates in asynchronous video interviews. As in most classical learning-based studies, the structured interviews are manually annotated by human experts on given rubrics.

More recent studies are focusing on question-answer pairs using hierarchical models based on neural networks. For example, Hemamou et al. [25] suggest a new hierarchical attention neural network, which aims to automatically extract verbal and non-verbal behaviors in asynchronous video interviews and classify candidates into hireable and non-hireable classes. The network model is able to represent the hierarchical and sequential structure of an interview assessment. In other words, it considers the sequentiality of the multimodal features present in the interview, as well as the hierarchical structure of an interview involving the candidate, the answer, and the word levels. A fine-grained interview-assessment approach based on long short-term memory and a hierarchical keyword-question attention mechanism is proposed by C. Chen et al. [6]. The method automatically assesses the candidates' personality traits and predicts their overall interview scores. Most recently, work by K. Chen et al. [29] aimed to automatically assess the competency of candidates based on textual features obtained through ASR. To do this, the study developed a hierarchical reasoning graph neural network model. The model allows consideration of the dependency relation and semantic-level interaction between questions and answers.

The multi-criteria multimodal nature of the problem causes much of the automated assessment and selection approaches to suffer from highly complex modeling and lack of large and adequate datasets (see [19,20]). The effectiveness of automated assessment of video interviews depends on factors such as unbiased data and annotation, effective feature extraction, integration of information from multiple criteria and modalities, and learning algorithms [20,21]. In addition, models using sophisticated learning algorithms such as deep learning are not able to explain the reasons for candidates' results. Therefore, how to derive, represent, weight, combine, and rank the outcomes of automated assessments becomes a critical issue for developing an effective, practical and explainable approach. Integrating MCDM techniques, especially for weighting, aggregation and ranking issues, would be a promising approach, which is also one of the major concerns of this study.

The second research direction involves studies that treat the assessment and selection process as an MCDM problem. A common characteristic of these studies is that the suitability of candidates is evaluated by interviewers with respect to subjective criteria in face-to-face interviews and/or by test-oriented objective approaches. None of these studies reports modeling of automated assessments of asynchronous video interviews. Many different MCDM methods (summarized below) have been proposed for personnel selection. In particular, most recent studies suggest hybrid approaches where typically one method is used for criteria weighting, and another one for evaluating candidates based on these criteria. More sophisticated approaches in which more than two methods are combined are also available (e.g., [30,31]). Another common characteristic of recent studies is that various extensions of the MCDM methods based on linguistic, fuzzy, or grey values are introduced to deal with the uncertainty inherent in the assessment process. Unlike many of the automated assessment frameworks discussed in the first research direction, MCDM approaches provide insight into the decision-making rationale used to draw a conclusion.

Here, only the most recent MCDM studies will be summarized. For more comprehensive reviews, the reader should refer to Chuang et al. [32], Kilic et al. [33], and Yalçın and Yapıcı Pehlivan [34]. Karabašević et al. [35] propose a hybrid MCDM approach for personnel selection based on the methods SWARA and ARAS. SWARA is preferred for

weighting the criteria, since it requires less pairwise comparison compared to other methods such as AHP; and ARAS is used for the evaluation of the candidates. In another study by Karabasevic et al. [36], SWARA and EDAS are used for weighting and evaluation purposes, respectively. Interviewers often prefer to use different linguistic terms to evaluate candidates, for which reason Liu et al. [37] suggest an extended VIKOR method based on interval 2-tuple linguistic variables for choosing the best candidates. In another study, Sang et al. [38] present an improved fuzzy TOPSIS method using the Karnik–Mendel algorithm to solve the personnel selection problem under uncertain information. Ji et al. [39] develop a projection-based TODIM method based on multi-valued neutrosophic numbers to be able to handle the hesitancy and fuzziness in the processes of personnel selection. Luo and Xing [31] propose a hybrid decision-making framework where they integrate PROMETHEE into MABAC to overcome the compensatory assumption of the MABAC method, since non-compensation among some criteria may occur in personnel selection problems. They also use an extended Best–Worst Method based on linguistic values to weight criteria. Özgörmüş et al. [40] distinguish between social and technical criteria in personnel selection. They propose an integrated fuzzy QFD-MCDM framework where Fuzzy DEMATEL is used to weight social criteria and Fuzzy QFD for technical criteria. Then, Fuzzy Grey Relational Analysis is employed to rank candidates by considering these criteria weights. Yalçın and Yapıcı Pehlivan [34] present a methodology where Hesitant Fuzzy Linguistic Term Sets based on comparative linguistic expressions are used to evaluate criteria and candidates in the personnel selection problem. Candidates are finally ranked by applying a fuzzy extension of the CODAS method. Krishankumar et al. [41] used intuitionistic fuzzy sets (IFSs) to represent the judgments of experts and integrated the VIKOR method with IFS to effectively deal with the personnel selection problem. To handle data that are vague and grey, Ulutaş et al. [42] suggested a grey extension of the PIPRECIA method for the evaluation of criteria importance. For the final ranking of the candidates, they used a grey extension of OCRA-G. Another hybrid MCDM approach proposed by Chuang et al. [32] applied rough set theory to derive the degree of interdependence and significance relation among criteria, which were merged into a relation matrix. They used a DEMATEL-based analytical network process to derive weights of criteria from this matrix. Next, the PROMETHEE-AS method was used to determine the ranking of candidates. Finally, C.-T. Chen and Hung [30] suggested a two-phase model for personnel selection that integrates TOPSIS, entropy method, 2-tuple linguistic variables, and PROMETHEE.

Consequently, a new approach is required that combines the benefits of learning-based techniques with MCDM techniques. This approach should be completely data-driven and efficient in order to minimize expert intervention and the subjectivity inherent in the assessment process. A flexible aggregation method that is able to aggregate assessment scores obtained in different forms and scales using objectively determined weights is needed. The approach should be able to consider (i) the interrelationships between (sub-)criteria; (ii) uncertainty inherent in the assessment process; and (iii) the accuracy levels of the learning algorithms. Additionally, it should fulfill all these requirements effectively and efficiently for a large number of (sub-)criteria and candidates. To address these requirements, this paper suggests an extended CBD approach to personnel selection based on automated AVI assessment scores.

3. The Theoretical Basis of the Proposed Approach

In order to clarify the theoretical foundations of the suggested approach, this section introduces the formal definitions and properties of the methods used in the approach.

3.1. Cumulative Belief Degree Approach

Introduced by Kabak and Ruan [23], the CBD approach aims to represent any information by a belief structure based on fuzzy linguistic terms. Before the formal definition of belief structure is given, it will be useful to define a fuzzy linguistic term set. Consider a finite and completely ordered linguistic term set $S = \{s_k\}$, $k \in \{0, \dots, K\}$, where s_k

denotes a possible value for a linguistic variable. The semantics of the linguistic terms in S is given by fuzzy numbers defined in the interval $[0, 1]$, which are characterized by their membership functions [43]. For instance, a set of five linguistic terms and their meanings could be described as follows:

$S = \{s_1 : \text{very poor}, s_2 : \text{poor}, s_3 : \text{fair}, s_4 : \text{good}, s_5 : \text{very good}\}$. Note that sets of linguistic terms may differ according to the nature of the problem [44].

Definition 1. The belief structure as used in this study can be defined as follows (cf. [23]):

$$B_{ij} = \left\{ \left(\beta_{ijk}, s_k \right), k = 1, \dots, k \right\}, \forall i, \forall j \quad (1)$$

$$\sum_{k=1}^k \beta_{ijk} \leq 1, \forall i, \forall j \quad (2)$$

where i and j denote the alternatives and criteria, respectively, and β_{ijk} denotes the belief degree for i th alternative with respect to the j th criterion at s_k level.

The belief structure is used to represent a distribution of belief degrees over a set of linguistic terms, regarding the fulfillment of a criterion by an alternative. For example, an alternative's ($i = 1$) performance score on a particular criterion ($j = 1$) evaluated by an expert is associated with linguistic term s_2 (poor) with 70% confidence and linguistic term s_3 (fair) with 30% confidence. Here, the belief degrees 70% and 30% indicate the extent to which the corresponding linguistic terms s_2 and s_3 are realized. The belief structure of this assessment can be expressed as follows: $B_{11} = \{(0.7, s_2), (0.3, s_3)\}$. The belief degrees corresponding to the other linguistic terms (i.e., s_1, s_4, s_5) are zero, and therefore are not shown.

One important distinguishing feature of the CBD approach is its ability to handle assessments in different forms without any loss of information. For example, assessments in the form of numerical, interval, linguistic, or fuzzy values can be easily transformed into belief structures. For details of transformation formulas suggested in the literature, the reader should refer to Ervural and Kabak [45]. The flexibility of belief structures in representing information also allows effective handling of missing values caused by lack of information or expert knowledge [23]. Another important feature of the CBD approach is its ability to aggregate values under vague and uncertain information environment.

To make operations on belief structures possible, Kabak and Ruan [23] suggest converting belief degrees into CBDs. This allows aggregating assessments at different linguistic term levels under multiple criteria.

Definition 2. CBD at s_k level can be defined as the aggregated belief degrees of the linguistic terms greater than or equal to s_k . More formally, CBD can be formulated as follows (cf. [23,46]):

$$C_{ij} = \left\{ \left(\gamma_{ijk}, s_k \right), k = 1, \dots, K \right\}, \forall i, \forall j \quad (3)$$

$$\gamma_{ijk} = \sum_{p=k}^K \beta_{ijp}, \forall i, \forall j \quad (4)$$

where γ_{ijk} is the CBD for the i th alternative with respect to the j th criterion at s_k level.

For instance, the cumulative belief structure of the belief degrees given in the example above can be formed as $C_{11} = \{(1, s_1), (1, s_2), (0.3, s_3), (0, s_4), (0, s_5)\}$. After CBDs are calculated for each criterion by using Equations (3) and (4), they are aggregated to obtain a final result, which indicates the total performance of an alternative.

Definition 3. A general formulation of the multi-criteria aggregation function is given below [47]:

$$C_i = \{(\gamma_{ik}, s_k), k = 1, \dots, k\}, \forall i \quad (5)$$

$$\gamma_{ik} = \sum_{j=1}^J w_j \gamma_{ijk} \quad (6)$$

where w_j denotes the weight (i.e., importance) of the j th criterion such that $w_j \in [0, 1]; j = 1, 2, \dots, J$; $\sum_{j=1}^J w_j = 1$; and γ_{ik} is the total performance score (i.e., aggregated CBD) of alternative i at level s_k .

One critical decision in the aggregation process is how to derive the weights of the criteria [48]. Various weighting methods are suggested in the literature, which can be grouped under subjective and objective methods. Direct rating and pairwise comparison are among the most common methods used for subjective weighting in the CBD literature (see [45,47,49,50]). Determining weights based on expert judgments is often a highly subjective, costly, and time-consuming task. Moreover, subjective judgments obtained from experts for the same criteria may significantly differ because of the difference in personal views, expertise, and backgrounds [51]. On the other hand, to obtain reliable weights, objective methods can be preferred. One of the most common methods used for objective weighting is the Ordered Weighted Averaging (OWA) operator (see [46,48,52]). Proposed by Yager [53], the OWA operator is a mean type aggregation function that associates weights with the ranks of the given values. Another approach proposed by Kabak and Ruan [23] aggregates values using rules, which are related to the existence of criteria.

Representing the results in form of a distribution of aggregated belief degrees over a set of linguistic terms provides more insight into the performance of an alternative to decision makers. This is another main benefit of the CBD approach. For ranking purposes, a single performance value may also be derived based on the final belief structure. The details of the ranking approach are provided in the next section.

3.2. Weighting of Criteria

To aggregate belief degrees in cumulative form, this paper suggests a weighted Bonferroni-OWA operator. This subsection introduces the formal definitions and properties of this operator.

3.2.1. Bonferroni Mean (BM) Operator

The choice of an appropriate aggregation function in multi-criteria problems is a critical issue. While standard operators such as averaging often yield reasonable results, more advanced aggregation functions are required to capture any existing interrelationship of criteria [24,54]. Originally proposed by [55], the Bonferroni mean is such an advanced mean type aggregation operator that is able to represent interrelationships between criteria. Because of its decomposable structure and distinguishable components, BM has been successfully applied to a wide range of problems [56–58]. The BM operator in its original form is formulated as follows.

Definition 4 ([55]). Let p and $q \geq 0$ and $\mathbf{x} = (x_1, \dots, x_n)$ be a vector of values where $x_i \in [0, 1]$, then the BM of these values is defined as:

$$B^{p,q}(\mathbf{x}) = \left(\frac{1}{n(n-1)} \sum_{i,j=1, i \neq j}^n x_i^p x_j^q \right)^{\frac{1}{p+q}} \quad (7)$$

Monotone and bounded by the min and max operators, the BM averages all the product pairs of non-identical inputs [54]. It is interpreted by Yager [24] as “a kind of combined averaging and anding operator”. The special case where $p = q = 1$ is commonly used to

aggregate values in multi-criteria problems. By rearranging the terms in Equations (7) and letting $p = q = 1$, the BM reduces to the following expression [24]:

$$B(\mathbf{x}) = \left(\frac{1}{n} \sum_{i=1}^n x_i u_i \right)^{\frac{1}{2}} \quad (8)$$

where x_i represents the performance score of alternative a with respect to criteria X_i ; $u_i = \frac{1}{n-1} \sum_{j=1, j \neq i}^n x_j$, $i = 1, \dots, n$, represents the average performance score concerning all criteria except X_i . In other words, each argument of the outer arithmetic mean in BM is the product of criterion score x_i with the average score of the remaining criteria x_j , $j \neq i$ [59].

3.2.2. Bonferroni Mean with OWA

To improve the modeling capability of BM, Yager [24] suggested some generalizations that involve replacing the simple average with other mean type aggregation functions such as OWA, as well as incorporating importance weights of criteria. Here, the OWA extension of BM will be explained. As mentioned above, the OWA operator does not associate weights directly with criteria, but rather with the ranks of scores on these criteria.

Definition 5 ([24]). Let $\mathbf{x}_{j \neq i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ denote a vector in $[0, 1]^{n-1}$ and \mathbf{w} denote an OWA weighting vector of dimension $n - 1$ with components $w_k \in [0, 1]$ when $\sum_k w_k = 1$. Then, the OWA function of $\mathbf{x}_{j \neq i}$, $j \neq i$ that replaces the inner arithmetic mean u_i in BM can be defined as follows:

$$OWA_{\mathbf{w}}(\mathbf{x}_{j \neq i}) = \sum_{k=1}^{n-1} w_k x_{(k)} \quad (9)$$

where the $(.)$ notation denotes the components of $\mathbf{x}_{j \neq i}$ being arranged in non-increasing order $x_{(1)} \geq x_{(2)} \geq \dots \geq x_{(n-1)}$.

Using this function, the extended BM can be expressed as follows [24]:

$$BON - OWA(\mathbf{x}) = \left(\frac{1}{n} \sum_{i=1}^n x_i OWA_{\mathbf{w}}(\mathbf{x}_{j \neq i}) \right)^{\frac{1}{2}} \quad (10)$$

In order to perform the $BON - OWA(\mathbf{x})$ operator, the weighting vector of dimension $n - 1$ needs to be specified. Based on this vector, the OWA operator is able to provide a rich family of aggregation functions varying from an AND operator (satisfying all the criteria) to an OR operator (satisfying at least one of the criteria) [53]. In other words, the form of the weighting vector determines the nature of the aggregation and therefore needs to be carefully chosen. Several approaches have been proposed for this purpose, which involve different techniques such as linguistic quantifiers [53], orness and entropy [60,61], exponential smoothing [62], linear objective-programming [63], normal distribution [64], kernel density estimation [65], and convolutional neural networks [66].

In this study, a method based on a measure of dispersion suggested by O'Hagan [60] and analytically solved by Fuller and Majlender [61] is used. This method neither needs to specify a particular monotone continuous function nor use any empirical data. To characterize the behavior of the OWA operator, the measures of orness and dispersion of the aggregation need to be calculated. The measure of orness introduced by Yager [53] is defined as follows.

Definition 6 ([53]). Let $\mathbf{w} = (w_1, \dots, w_n)$ indicate the OWA weighting vector. The degree of orness associated with the weighting vector of OWA is defined as:

$$orness(\mathbf{w}) = \frac{1}{n-1} \sum_{i=1}^n (n-i)w_i \quad (11)$$

where $orness(\mathbf{w}) = \alpha$ is a situation parameter and characterizes the degree to which the aggregation is like an OR operation. Note that for any weighting vector, $orness(\mathbf{w}) \in [0, 1]$.

The concept of dispersion, as a kind of entropy, measures how much of the information in the arguments included in the aggregation is really used. It is formally defined as follows.

Definition 7. The measure of dispersion of \mathbf{w} is formulated as

$$dispersion(\mathbf{w}) = - \sum_{i=1}^n w_i \ln w_i \quad (12)$$

The more uniformly the weights are distributed, the more information is considered in the aggregation. The maximum value of dispersion is obtained with a \mathbf{w} where $w_i = 1/n$ for every $i = 1, \dots, n$ [67].

O'Hagan [60] formulates the OWA weights determination problem as a constrained non-linear optimization model where dispersion (i.e., objective function) is maximized for a predefined degree of orness (i.e., constraint). The optimization model suggested by O'Hagan [60] is as follows:

$$\text{Maximize : } - \sum_{i=1}^n w_i \ln w_i \quad (13)$$

$$\text{Subject to : } \frac{1}{n-1} \sum_{i=1}^n (n-i)w_i = \alpha, \quad 0 \leq \alpha \leq 1, \quad (14)$$

$$\sum_{i=1}^n w_i = 1, \quad 0 \leq w_i \leq 1, \quad i = 1, \dots, n \quad (15)$$

Fuller and Majlender [61] applied the method of Lagrange multipliers to convert the constrained optimization problem to a polynomial equation that can be solved analytically to derive the optimal weighting vector. The weighting vector is obtained by solving the following equations:

$$w_j = \sqrt[n-1]{w_1^{n-j} w_n^{j-1}} \quad (16)$$

$$w_n = \frac{((n-1)\alpha - n)w_1 + 1}{(n-1)\alpha + 1 - nw_1} \quad (17)$$

$$w_1[(n-1)\alpha + 1 - nw_1]^n = ((n-1)\alpha)^{n-1} [((n-1)\alpha - n)w_1 + 1] \quad (18)$$

First, the value of w_1 is obtained for a given orness degree, using Equation (18). Then, by substituting the value of w_1 in Equation (17) w_n is determined. After w_1 and w_n are obtained, all other weights can be calculated using Equation (16). Note that (i) if $n = 2$, then $w_1 = \alpha$ and $w_2 = 1 - \alpha$; (ii) if $\alpha = 0$ or $\alpha = 1$, then the weighting vectors are obtained as $\mathbf{w} = (0, 0, \dots, 1)$ and $\mathbf{w} = (1, 0, \dots, 0)$, respectively, with $dispersion(\mathbf{w}) = 0$; (iii) if $n \geq 3$ and $0 < \alpha < 1$, then Equations (16)–(18) are used to obtain the weighting vector [61].

4. Proposed Approach

In this subsection, the details of the weighted Bon-OWA-based CBD approach proposed for personnel selection are presented. It is explained how outcomes of automated AVI assessments of candidates can be derived, represented, weighted, combined, and ranked under multiple criteria. Extending the CBD approach by replacing the simple average with the weighted Bon-OWA operator improves its modeling capability such that it enables considering (i) accuracy levels of the learning algorithms as weights; (ii) inter-relationships between sub-criteria; and (iii) objective weights of (sub-)criteria. Thus, this completely data-driven and highly flexible approach effectively handles the complexity and uncertainty involved in the assessment and selection process and increases the capacity for handling large numbers of candidates.

The proposed approach consists of two consecutive stages: (i) the Assessments Stage, where patterns in candidate video responses are learned and predicted using AVI in combination with statistical and AI-based approaches; and (ii) the Selection Stage, where automated AVI assessment scores are transformed and aggregated using an extended CBD approach in a multi-criteria environment. The proposed approach combines the benefits of learning-based techniques with MCDM techniques and makes unique methodological contributions to the selection stage. The proposed approach, depicted in Figure 1, comprises the following steps.

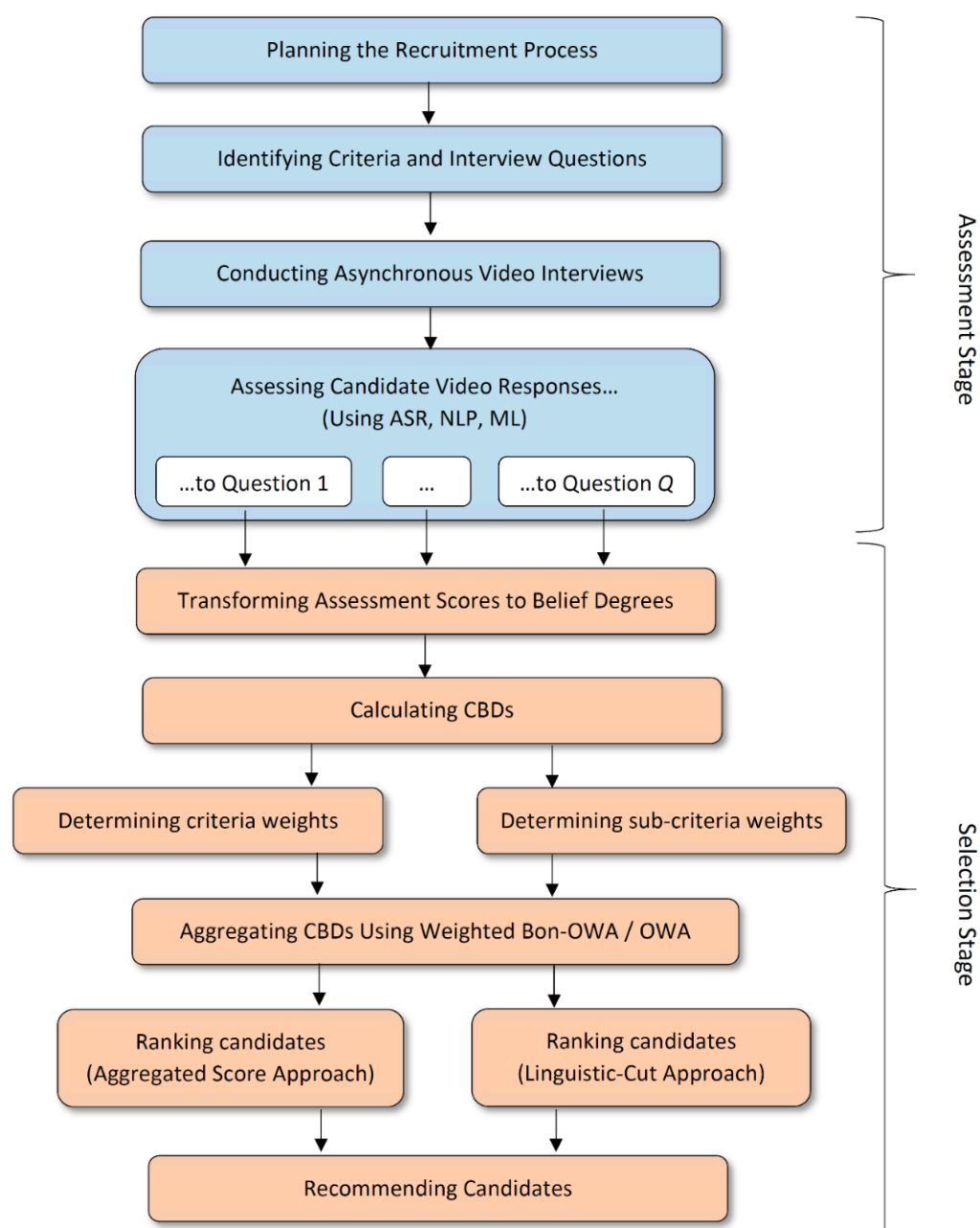


Figure 1. Steps of the proposed approach.

Step 1. Planning the Recruitment Process: The process begins with identifying job vacancies followed by preparing job descriptions that are required to describe the positions. The aim is to attract as many high-quality candidates as possible that should “fit” open positions. Job openings are advertised internally and externally on the company’s own

career sites and/or on popular social sites or job portals. Online advertising and searching is still the most preferred means to attract job seekers to the vacancies [4].

Step 2. Identifying Personnel Selection Criteria and Interview Questions: The next stage of the process aims to identify specifications that are required to fill the respective positions. Specifications indicate the hiring criteria used to assess candidates, which include knowledge, experience, specific skills and competencies, and personality traits, among others. These criteria are often evaluated through multiple questions. Once the assessment criteria are identified, the questions associated with each criterion to be asked in the asynchronous video interviews are formulated.

Step 3. Obtaining and Assessing Candidate Video Responses: Candidates who wish to apply for the open position(s) are first requested to register online and send their resumes. After automatically screening candidates' available information, only candidates that match the basic requirements (e.g., educational qualification, work experience) necessary to proceed to the next stages of the process are determined. These selected candidates are invited to complete an online asynchronous video interview at any time and place convenient to them within a predefined time interval. In the interview, candidates are requested to record their responses to the predetermined set of interview questions and submit them online.

For each recorded response, verbal and/or non-verbal indicators of a candidate are extracted by using AI-based approaches to learn and predict patterns in the video response. For verbal content, the candidate responses are first converted to text using ASR and then mapped to feature vectors using NLP techniques (e.g., document embedding). For non-verbal cues, features such as tone of voice, facial expressions, and body gestures can be extracted using visual words and audio words. These visual/audio words are represented as feature vectors using NLP tools. Then, these vectors are used as input to ML algorithms to predict the interview performance with respect to each interview question. Predictions are obtained either as discrete class labels (and their probabilities, if available) through classification algorithms or as continuous values through regression algorithms. These predictions need to be combined effectively to assess a candidate's overall suitability for the position and the organization (see steps 4 and 5).

Step 4. Representing Automated AVI Assessment Scores with (Cumulative) Belief Degrees: Before making a decision on the most suitable candidate(s), the hiring managers would consider scores predicted for each single question and, more essentially, the overall suitability of candidates obtained by aggregating these scores. While this appears deceptively simple, there are a variety of issues that need to be addressed in the aggregation process, which can be summarized as follows. Based on the features extracted and the learning algorithms used, the predicted scores (or class information) can be in different forms and scales such as probability values, interval values, ratings, test scores, distances, or linguistic values. Thus, there is a need for a flexible and easy-to-use approach that enables the aggregation of scores in different forms. Moreover, the automated assessment scores may involve some uncertainty. Here, uncertainty is used in a broader sense that considers subjectivity, missing values, and inaccuracy as important sources. Essentially, automated assessments are based on learned patterns obtained from previous evaluations of interviewers and may carry some subjectivity. Missing data and accuracy of the learning algorithms are further factors that influence the aggregation of the scores and, in turn, the job suitability of candidates. Finally, the questions associated with each criterion used to assess candidates' video responses may be highly interrelated. To address these issues, an aggregation process based on CBDs is suggested and employed as follows.

Step 4.1. Transforming Scores to Belief Degrees: To be able to aggregate automated AVI assessment scores represented in different forms and scales, the scores are first transformed into belief structures as defined in Equations (1) and (2). The belief structure represents the fulfillment of a (sub-)criterion by a candidate through a distribution of belief degrees over a set of linguistic terms. The set of linguistic terms used in this study and their meanings are

described as follows $S = \{s_k\}, k \in \{1, \dots, 5\}$ where s_1 : *unsatisfactory*, s_2 : *needs improvement*, s_3 : *meets expectations*, s_4 : *exceeds expectations*, s_5 : *exceptional*.

The method of associating a candidate's score on a particular question with linguistic terms depends on the form and scale of the assessment. Automated AVI assessment scores are often obtained in form of probability values, ratings, test scores, or distances. For these types of values, the direct value assignment approach suggested by Kabak and Ruan [23] can be applied to transform the values into belief structures. According to this approach, scores based on predefined scales (e.g., 0–100) are transformed into belief degrees through membership functions ($\mu_{\tilde{s}_k}$) defined for each linguistic term (s_k) and interview question (q). More formally, the belief structure related to a given score v_{ijq} is formulated as follows [23]:

$$B(v_{ijq}) = \left\{ \left(\mu_{\tilde{s}_k}(v_{ijq}), s_k \right), k = 1, \dots, K \right\}, \quad \forall i, \forall j, \forall q \quad (19)$$

The membership function of each linguistic term is defined as a triangular fuzzy number (TFN). A TFN of linguistic term s_k is simply designated with three parameters (l_k, m_k, u_k) and its membership function $\mu_{\tilde{s}_k}(v_{ijq})$ is defined as [68]:

$$\mu_{\tilde{s}_k}(v_{ijq}) = \begin{cases} (v_{ijq} - l_k) / (m_k - l_k), & l_k \leq v_{ijq} \leq m_k \\ (u_k - v_{ijq}) / (u_k - m_k), & m_k \leq v_{ijq} \leq u_k \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

where l_k and u_k denote the lower and upper values of the support of TFN of linguistic term s_k , respectively, and m_k denotes the modal value. The linguistic term set used in this study consists of five terms; hence, for each term, a TFN has to be defined. Figure 2 shows an example for the transformation of a score ($v_{ijq} = 0.58$) in the form of probability (obtained from a classification algorithm) into belief degrees using Equation (20). The membership degrees 0.68 and 0.32 indicate the extent to which the corresponding linguistic terms s_3 and s_4 are realized, respectively. The belief structure of the given score will be formulated as $B_{ijq} = \{(0.68, s_3), (0.32, s_4)\}$.

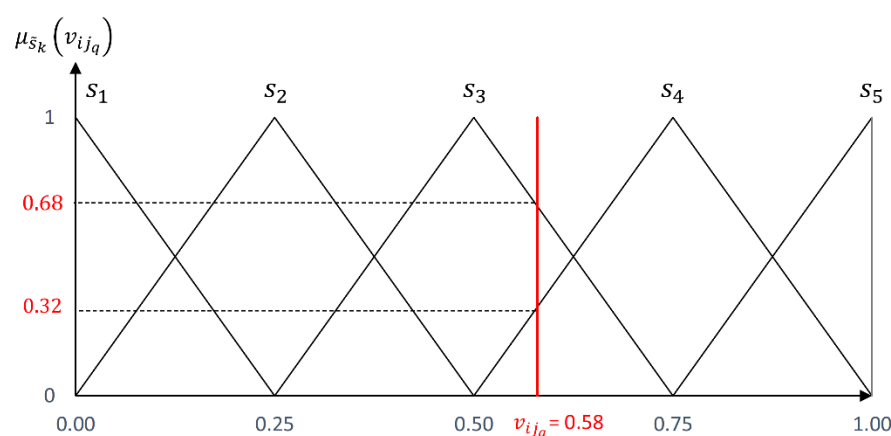


Figure 2. Transformation of a score into belief degrees.

In case the criteria are not limited to any specific range, the scores need to be first normalized before they are transformed into belief structures. Various normalization techniques such as linear normalization (e.g., min–max scaling, Tchebychev distance), vector normalization (e.g., scaling to unit length), non-linear normalization (e.g., sigmoid function, hyperbolic tangent) are available for this purpose.

As mentioned earlier, belief structures allow us to effectively deal with missing values in representing information. For example, if a candidate i does not/cannot respond to an interview question q , then the belief degrees β_{ijqk} of all linguistic terms except s_1 (unsatis-

factory) are set to zero. On the other hand, if the AI model technically fails to produce a score for the response of a candidate to a particular question, then all the linguistic term options will be considered as possible and the total belief is distributed to all linguistic terms evenly (see [23]).

Notice that besides the described direct value assignment approach, various transformation formulas have been suggested in the literature to handle other additional types of data (for more detail see [45]).

Step 4.2. Calculating Cumulative Belief Degrees (for each Interview Question): To make operations on belief structures possible, the belief degrees are converted into CBDs. This is necessary since belief structures cannot be directly ranked and they involve multiple dimensions rather than a single score [48]. Using Equations (3) and (4), the CBD ($\gamma_{ij,qk}$) for the i th candidate with respect to the q th interview question associated with the j th criterion at s_k level is calculated.

To explain the idea of CBDs, suppose that the suitability of a candidate is determined according to a threshold that is specified as one of the terms (e.g., s_3) in the linguistic term set. Then, the belief degrees of the terms that are greater than or equal to the threshold (i.e., s_3, s_4, s_5) give the total belief of the suitability of the candidate.

Step 5. Calculating Aggregated Scores: Since each criterion is evaluated through multiple interview questions that constitute a hierarchic composition, the cumulative belief structures are aggregated in a hierarchical way. First, CBDs obtained for interview questions associated with a particular criterion are aggregated using an objective weighting approach. Then, the resulting aggregated cumulative belief structures, each corresponding to a particular criterion, are aggregated to determine a candidate's overall suitability for the position.

Step 5.1. Determining Criteria and Sub-Criteria Weights: To avoid highly subjective, costly, and time-consuming tasks and ensure a completely data-driven assessment and selection process without the need for any expert intervention, the OWA operator is used. The optimal OWA weighting vector (\mathbf{w}) is obtained by applying Equations (16)–(18). As explained in Section 3, the behavior of the OWA operator depends on a predefined level of orness (α). According to the chosen level, two alternative weighting schemes can be distinguished for the personnel selection problem:

- A high orness degree is chosen to place more importance on higher scores without ignoring lower ones. This avoids missing any promising candidate who has (very) low scores on only few (sub-)criteria.
- A low orness degree is chosen to place more importance on lower scores without ignoring higher ones. This avoids considering candidates with unacceptably low scores on some (sub-)criteria.

Step 5.2. Aggregating CBDs for each candidate at question and criteria level: For each candidate, the CBDs obtained for interview questions associated with a particular criterion are aggregated using an extended version of the BON-OWA Operator. The proposed operator incorporates the accuracy of a learning algorithm as a weighting factor. More accurate models are assigned higher weights to make sure that these models have a greater effect on the final result. Several different accuracy measures are suggested in the literature. For example, proportion of correct classifications, area under the curve (AUC), and F1 score are common accuracy measures used in classification problems. On the other hand, the accuracy of a regression model can be represented by the root mean squared error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (R^2), among others.

Definition 8. Given the accuracy ($a_{j_q} \in [0, 1]$) of the learning algorithms formulated for the questions associated with criterion j , the suggested Weighted BON-OWA operator can be formulated as follows:

$$\gamma_{ijk} = \left(\frac{1}{A} \sum_{q=1}^Q a_{j_q} \gamma_{ij_q k} \text{OWA}_w(\gamma_{ij_{p \neq q} k}) \right)^{\frac{1}{2}}, \quad \forall i, \forall j, \forall k \quad (21)$$

$$\text{OWA}_w(\gamma_{ij_{p \neq q} k}) = \sum_{r=1}^{Q-1} w_{j_r} \gamma_{ij_{(r)} k} \quad (22)$$

where $\gamma_{ij_{p \neq q} k} = (\gamma_{ij_1 k}, \dots, \gamma_{ij_{q-1} k}, \gamma_{ij_{q+1} k}, \dots, \gamma_{ij_Q k})$ denotes a vector in $[0, 1]^{Q-1}$, $A = \sum_{q=1}^Q a_{j_q}$, w denotes an OWA weighting vector of dimension $Q-1$ with components $w_{j_r} \in [0, 1]$ when $\sum_r w_{j_r} = 1$. The (\cdot) notation denotes the components of $\gamma_{ij_{p \neq q} k}$ being arranged in non-increasing order $\gamma_{ij_{(1)} k} \geq \gamma_{ij_{(2)} k} \geq \dots \geq \gamma_{ij_{(Q-1)} k}$.

This aggregation operator is monotonic concerning the arguments and this satisfies the boundary conditions when all $\gamma_{ij_q k} = 1$ or $\gamma_{ij_q k} = 0$. Notice that when $a_{j_q} = 1$ for all q , then the original case is obtained. At least two non-zero inputs (i.e., $\gamma_{ij_q k}, \gamma_{ij_p k} > 0$) are required to obtain a non-zero output. To avoid a zero score for the case of only one non-zero input, the arithmetic mean can be performed for the particular pair.

Once the aggregated cumulative belief structures are obtained for each criterion, the next step is to aggregate these values to determine a candidate's overall suitability for the position. To do this (after ranking γ_{ijk} values in descending order), the simple average in Equation (6) is replaced by the classical OWA function as shown below:

$$\gamma_{ik} = \sum_{r=1}^I w_r \gamma_{i(r)k}, \quad \forall i, \forall k \quad (23)$$

Step 6. Ranking Candidates: Two approaches can be used to compare and rank the final cumulative belief structures of the candidates.

- *Ranking candidates using the Aggregated Score Approach:* The candidates are ranked according to the overall suitability scores obtained by converting the final cumulative belief structures into single values. This is achieved by first assigning predefined numerical values to the linguistic terms, then decomposing CBDs into belief degrees, and finally summing the product of each numeric value and its associated belief degree. More formally, the overall suitability score (OSS) of each candidate is calculated by the following formula [23]:

$$\text{OSS}_i = \left(\sum_{k=1}^{K-1} v_k (\gamma_{ik} - \gamma_{ik+1}) \right) + v_K \gamma_{iK}, \quad \forall i \quad (24)$$

where v_k denotes a numerical value for the linguistic term s_k and γ_{ik} denotes the overall suitability of candidate i at level s_k . If no other information is available, the numerical values of each linguistic term are set to $v_k = k$, $k = 1, \dots, K$. To express OSSs in percentage, the scores should be multiplied by $100/K$ [48].

- *Ranking Candidates Using the Linguistic-Cut Approach:* According to this approach, the final cumulative belief structure of a candidate is transformed into a single linguistic term. First, a threshold (τ) is chosen that specifies the minimum CBD required to sufficiently fulfill a linguistic term. Accordingly, a single term s_k is assigned to a candidate if s_k is the highest linguistic term with a CBD greater than or equal to τ . It is also possible to determine the threshold by examining a graphical representation of the results [23]. By considering alternative thresholds, sensitivity and robustness

analyses can be performed [45]. The transformation into a single linguistic term can be defined as follows [52]:

$$LSS_i^\tau = \text{Sup}_{k=1,\dots,K}[s_k | \gamma_{ik} \geq \tau] \quad (25)$$

where LSS_i^τ denotes the Linguistic Suitability Score of candidate i for a given threshold value $\tau \in [0, 1]$. Candidates are ranked based on their linguistic suitability scores. The hiring manager may set a minimum expectation (in terms of LSS) for the open position to be met by the candidates.

Step 7. Recommending Candidates: The list of recommended candidates and important insights are then shared with the hiring manager(s) for review. The hiring manager(s) review(s) the candidates' interviews, resumes, and suitability scores. Consequently, a final decision on the shortlist of candidates that will move through the next stages of the interview process is made. In this final step of the proposed assessment and selection process, the chosen candidates are invited for an online (synchronous) interview with the hiring manager(s) and/or HR executive.

5. Example Case

In this section, an example case is presented to illustrate the applicability and effectiveness of the proposed approach. In the example, a HR software company is in search of a sales specialist to join their company. The main concern of the company is to apply a completely data-driven approach in the early stages of the hiring process to screen for high-quality candidates to be invited for an initial in-person interview. This will minimize expert intervention and the subjectivity inherent in the assessment and selection process. The steps of the process are explained below.

Step 1. Planning the Recruitment Process: First, a team responsible for the hiring process was set up, including a hiring manager and a recruiter. The job description and qualifications required to fill the open position (sales specialist) were advertised online on popular job portals. To encourage maximum possible job seekers to apply and make the hiring process more effective and efficient, the team decided to use AVI. The suitability of the candidates was predicted using an automated video interview assessment process where only the verbal content of the recorded responses was analyzed. The interviews, assessments of video responses, and the selection (shortlisting) of potential candidates were planned to be completed in one month. According to the plan, the shortlist should not exceed three candidates who would advance to the next stages of the interview process.

Step 2. Identifying Personnel Selection Criteria and Interview Questions: In this step, the hiring criteria used to assess the candidates were established. The team agreed on using a competency-based assessment process, since using AVI in combination with AI-based approaches allows candidates' competencies to be assessed in earlier stages of the assessment and selection process. Although competencies can be more difficult to measure than hard skills, they provide more insight into a candidate's past experiences, abilities, and behavioral patterns. Every position is unique and requires a unique set of competencies. The competencies (i.e., criteria) that were regarded as essential for the open position were communication, persuasiveness, and results orientation. To get a broader view, each competence was evaluated through multiple behavioral questions. These interview questions focus on actions in particular workplace situations, which allow a candidate's experience to be compared to the requirements of the open position. The behavioral competencies and the questions associated with each competency to be asked in the asynchronous video interviews are summarized in Table 1.

Step 3. Obtaining and Assessing Candidate Video Responses: First, the resumes of the applicants who applied for the open position were automatically screened. Only 12 out of 19 applicants matched the basic requirements (educational qualification, work experience, etc.) necessary to proceed to the next stages of the process. The selected candidates were invited to complete an online asynchronous video interview at any time and place

convenient to them within one week. The online interview could be completed using a PC with an external or built-in webcam or mobile device with a built-in camera. In the asynchronous video interview, candidates were requested to record their responses to nine interview questions presented in written format (listed in Table 1) and submit them online. For each question, candidates were given one minute to read and organize their thoughts before recording began and four minutes maximum to respond. Candidates did not have the option to review their recorded responses or to re-record them; however, they were given the opportunity to practice on sample questions before starting the interview.

Table 1. Behavioral competencies and associated questions.

Competency (j)	Definition	Interview Question (j_p)
Communication ($j = 1$)	Considering and responding to the needs, ideas, and feelings of individuals or groups in a professional, clear, and accurate manner.	<p>Tell about an experience in which you explained a complex technical problem to a person who does not understand technical jargon? How did you handle this delicate situation? What was the result? (j_1)</p> <p>How do you avoid “verbal overkill”? How do you reduce messages to their essence without losing the main intent and content? Give an example. (j_2)</p> <p>Tell about a time when you had to explain something you knew well to someone who had difficulty understanding the subject. How did you do it and what was the result? (j_3)</p>
Persuasiveness ($j = 2$)	Using appropriate interpersonal styles and communication methods to persuade, influence, or impress others in order to achieve understanding and acceptance of a product, service, or idea.	<p>Tell about an experience in which you effectively conveyed your opinion to others. What was the idea you conveyed? What was the result? What would you do differently if you were faced with the same situation again? (j_1)</p> <p>Tell about an experience in which you did not give up and made a determined effort despite having difficulty in convincing the people around you. What were the reactions you encountered? What were the things that challenged you? What was the result? (j_2)</p> <p>What was the most stressful professional negotiation you have been involved in? How did you handle it? (j_3)</p>
Results Orientation ($j = 3$)	Focusing effort on desired outcomes and the ways to achieve them by translating ideas into concrete actions, removing barriers, and mobilizing resources.	<p>Tell about a situation where you had to take several actions over a period of time and overcome obstacles in order to achieve a business objective. What were the obstacles in your path? What was the result? (j_1)</p> <p>Tell about an experience in which you accomplished a job you pursued effectively and quickly. What did you do to complete the job as quickly and effectively as possible? (j_2)</p> <p>Describe a stretch goal or objective that you were able to achieve. Why was this a stretch goal? What was the result? (j_3)</p>

Once a candidate had finished the interview, the suitability of the candidate was predicted using an automated assessment process. More precisely, for each recorded response, the verbal content was scored automatically using AI-based approaches. The candidate responses were first converted to text using the Google Speech-to-Text API [69]. This API powered by Google’s AI technologies apply the most recent deep-learning neural-network algorithms for ASR.

After the interview responses were preprocessed, they were transformed into vectors using pretrained models. These vectors, also known as embeddings, represent the meaning of a text (or any other unit of natural language) as a point in a multidimensional vector space [70]. Texts that are closer in the semantic space are expected to be more similar in meaning. In the current example, document embedding based on fastText [71] was used, which is a library for learning of word embeddings developed by Facebook’s Artificial Intelligence Research Lab. Consequently, a vector was obtained for each response by aggregating word embeddings. This low-dimensional representation (consisting of 300 dimensions) allows us to simply integrate candidate responses in form of real-valued

vectors into ML models. Table 2 shows the vector representations of the responses given by Candidate 1.

Table 2. Vector representations of the responses of Candidate 1.

Competency	Response	Dim 1	Dim 2	...	Dim 299	Dim 300
Communication ($j = 1$)	Response to j_1	0.0078009	−0.0041856	...	−0.0122872	−0.0117441
	Response to j_2	0.0279355	−0.0020948	...	−0.0029779	−0.0140276
	Response to j_3	0.0081928	−0.0036274	...	−0.0012601	−0.0162252
Persuasiveness ($j = 2$)	Response to j_1	−0.0034173	−0.0187041	...	−0.0015710	−0.0123723
	Response to j_2	−0.0057934	−0.0288823	...	−0.0015282	−0.0124730
	Response to j_3	0.0022617	−0.0203922	...	−0.0067732	−0.0028028
Results Orientation ($j = 3$)	Response to j_1	0.0162339	−0.0035523	...	−0.0093317	0.0059282
	Response to j_2	0.0281698	0.0122494	...	0.0081218	0.0090014
	Response to j_3	0.0134955	0.0024308	...	0.0103870	0.0043022

The embeddings were used as input to a trained binary classification model to predict the interview performance of a candidate with respect to each interview question. The classifier, based on a multi-layer neural network model, predicts the probabilities with which a response of a candidate belongs to the following classes: “not suitable” or “suitable”. The choice for a binary classifier was motivated by the fact that a binary classifier is much less complicated than a multi-class classifier. More precisely, binary classifiers often produce more accurate results for even smaller datasets, but at a cost of a less detailed classification. The accuracy of each applied classification model and the predictions, obtained in terms of probabilities, is presented in Table 3. Each probability value indicates the likelihood of belonging to the class “suitable”. For example, the response of Candidate 1 to the first interview question of the competency “Communication” has a likelihood of 0.58 to be suitable. The accuracy of a classifier was computed using 10-fold cross validation.

Table 3. Predictions of candidates’ interview performance with respect to each interview question.

Candidate	Communication ($j = 1$)			Persuasiveness ($j = 2$)			Results Orientation ($j = 3$)		
	j_1	j_2	j_3	j_1	j_2	j_3	j_1	j_2	j_3
	(0.82) *	(0.74) *	(0.78) *	(0.79) *	(0.75) *	(0.84) *	(0.81) *	(0.82) *	(0.79) *
1	0.58	0.66	0.47	0.35	0.33	0.41	0.82	0.93	0.59
2	0.52	0.62	0.51	0.56	0.65	0.33	0.78	0.76	0.73
3	0.15	0.22	0.25	0.34	0.28	0.31	0.10	0.17	0.14
4	0.92	0.87	0.85	0.88	0.90	0.83	0.71	0.68	0.60
5	0.72	0.70	0.61	0.47	0.73	0.68	0.38	0.42	0.27
6	0.84	0.91	0.66	0.80	0.74	0.85	0.83	0.88	0.89
7	0.18	0.00	0.14	0.23	0.16	0.10	0.17	0.13	0.24
8	0.42	0.50	0.39	0.36	0.43	0.38	0.37	0.42	0.30
9	0.62	0.64	0.75	0.56	0.60	0.73	0.42	0.57	0.65
10	0.89	0.80	0.95	0.68	0.45	0.59	0.63	0.70	0.57
11	0.26	0.35	0.25	0.43	0.17	0.27	0.76	0.72	0.80
12	0.48	0.63	0.55	0.56	0.63	0.43	0.52	0.62	0.57

* Accuracy (F1 score) of the classification model

Step 4. *Representing Automated AVI Assessment Scores with (Cumulative) Belief Degrees:* To be able to compare the results for each single question and effectively combine them to assess a candidate’s overall suitability, the predictions obtained in step 3 were transformed into (cumulative) belief degrees.

Step 4.1. *Transforming Scores to Belief Degrees:* Based on the features extracted and the learning algorithms used, the predicted scores were obtained in the form of class memberships and associated probabilities (v_{ijq}). Appropriate for this type of value, the direct value assignment approach was employed to transform the probabilities into belief

structures. The set of fuzzy linguistic terms over which distributions of belief degrees are represented was formulated as $S = \{s_k\}, k \in \{1, \dots, 5\}$ where s_1 : *unsatisfactory*; s_2 : *needs improvement*; s_3 : *meets expectations*; s_4 : *exceeds expectations*; s_5 : *exceptional*. The membership function $\mu_{\tilde{s}_k}$ of each linguistic term s_k (given in Figure 2) was defined as a TFN using Equation (20). Table 4 presents the TFNs defined for the given linguistic terms.

Table 4. TFNs defined for the linguistic terms.

Linguistic Term	Linguistic Expression	TFN
s_1	<i>Unsatisfactory</i>	(0, 0, 0.25)
s_2	<i>Needs Improvement</i>	(0, 0.25, 0.50)
s_3	<i>Meets Expectations</i>	(0.25, 0.50, 0.75)
s_4	<i>Exceeds Expectations</i>	(0.50, 0.75, 1)
s_5	<i>Exceptional</i>	(0.75, 1, 1)

According to the direct value assignment approach, probabilities were transformed into belief degrees through the membership functions ($\mu_{\tilde{s}_k}$) defined for each linguistic term (s_k) and interview question (q) as shown in Equation (20). For example, the belief structure associated with the response of Candidate 1 to the first interview question of the competency “Communication” (i.e., for $v_{11_1} = 0.58$) is determined as follows:

$$\begin{aligned}\mu_{\tilde{s}_1}(0.58) &= \begin{cases} (0.25 - v_{ij_q}) / (0.25 - 0), & 0 \leq v_{ij_q} \leq 0.25 \\ 0, & \text{otherwise} \end{cases} \\ \mu_{\tilde{s}_2}(0.58) &= \begin{cases} (v_{ij_q} - 0) / (0.25 - 0), & 0 \leq v_{ij_q} \leq 0.25 \\ (0.50 - v_{ij_q}) / (0.50 - 0.25), & 0.25 \leq v_{ij_q} \leq 0.50 \\ 0, & \text{otherwise} \end{cases} \\ \mu_{\tilde{s}_3}(0.58) &= \begin{cases} (v_{ij_q} - 0.25) / (0.50 - 0.25), & 0.25 \leq v_{ij_q} \leq 0.50 \\ \frac{(0.58 - 0.25)}{(0.50 - 0.25)} = 0.68, & 0.50 \leq v_{ij_q} \leq 0.75 \\ 0, & \text{otherwise} \end{cases} \\ \mu_{\tilde{s}_4}(0.58) &= \begin{cases} \frac{(0.58 - 0.50)}{(0.75 - 0.50)} = 0.32, & 0.50 \leq v_{ij_q} \leq 0.75 \\ (1 - v_{ij_q}) / (1 - 0.75), & 0.75 \leq v_{ij_q} \leq 1 \\ 0, & \text{otherwise} \end{cases} \\ \mu_{\tilde{s}_5}(0.58) &= \begin{cases} (v_{ij_q} - 0.75) / (1 - 0.75), & 0.75 \leq v_{ij_q} \leq 1 \\ 0, & \text{otherwise} \end{cases}\end{aligned}$$

The membership degrees 0.68 and 0.32 indicate the extent to which the corresponding linguistic terms s_3 : *meets expectations* and s_4 : *exceeds expectations* are realized, respectively. Thus, the belief structure of the given probability is formulated as $B(v_{11_1}) = \{(0.68, s_3), (0.32, s_4)\}$. The belief structures associated with the responses of Candidate 1 to the remaining questions were calculated in a similar way. Table 5 presents the complete results for Candidate 1. For the belief structures of other candidates, the reader should refer to Appendix A. Note that one of the candidates ($i = 7$) did not respond to the second interview question of the competency “Communication”. For this particular question, the candidate’s belief degrees (β_{71_2k}) of all linguistic terms except s_1 were set to zero (see Table A6).

Table 5. Belief degrees of Candidate 1.

Competency	Question	Prediction	Belief Degrees					CBDs				
			s_1	s_2	s_3	s_4	s_5	s_1	s_2	s_3	s_4	s_5
Communication ($j = 1$)	j_1	0.58			0.68	0.32		1	1	1	0.32	0
	j_2	0.66			0.36	0.64		1	1	1	0.64	0
	j_3	0.47		0.12	0.88			1	1	0.88	0	0
Persuasiveness ($j = 2$)	j_1	0.35		0.6	0.4			1	1	0.40	0	0
	j_2	0.33		0.68	0.32			1	1	0.32	0	0
	j_3	0.41		0.36	0.64			1	1	0.64	0	0
Results Orientation ($j = 3$)	j_1	0.82				0.72	0.28	1	1	1	1	0.28
	j_2	0.93				0.28	0.72	1	1	1	1	0.72
	j_3	0.59			0.64	0.36		1	1	1	0.36	0

Step 4.2. Calculating Cumulative Belief Degrees (for each Interview Question): To make operations on belief structures and their ranking possible, the belief degrees were converted into CBDs (γ_{ijk}) using Equations (3) and (4). For instance, the CBD of Candidate 1 with respect to the first interview question associated with the competency “Communication” at level s_3 is calculated as follows:

$$\gamma_{11,3} = \sum_{k=3}^5 \beta_{11,k} = 0.68 + 0.32 + 0 = 1$$

The sum of the belief degrees of the terms that are greater than or equal to the specified level s_3 gives the cumulative belief on the suitability of the candidate at this level. Accordingly, the cumulative belief structure of Candidate 1 with respect to question 1 associated with the competency “Communication” can be formed as $C_{11,1} = \{(1, s_1), (1, s_2), (1, s_3), (0.32, s_4), (0, s_5)\}$. The CBDs associated with each response of Candidate 1 are given in Table 5. The results indicate that Candidate 1 just met expectations for the competency “Persuasiveness” but achieved a much better interview performance for the competency “Results Orientation” by exceeding expectations.

Step 5. Calculating Aggregated Scores: Since the assessment and selection process involved multiple competencies, each of which was evaluated through multiple interview questions, a candidate’s overall suitability for the open position was determined by aggregating the cumulative belief structures in a hierarchical way. To ensure a completely data-driven process without the need for any expert intervention, the aggregations at both question and competency levels were performed by objective weighting approaches.

Step 5.1. Determining Weights: Before aggregating CBDs (at question level) with the proposed Weighted BON-OWA operator, the optimal OWA weighting vector (\mathbf{w}_j) was determined by applying Equations (16)–(18). As the form of the weighting vector depends on the level of orness (α) and the hiring manager did not want to miss promising candidates who had low scores on only a few questions, a high orness degree was chosen to place more importance on higher scores. So, the level of orness was set to 0.7, which lies between the average and max operators. The Weighted BON-OWA operator requires $n_j = Q - 1$ weights to be determined, where Q denotes the total number of questions associated with a particular competency j . Accordingly, the optimal two-dimensional OWA weighting vector was determined as $\mathbf{w}_j = (0.7, 0.3)$. Note that $n_j = 2$ is a special case where $w_{j_1} = \alpha$ and $w_{j_2} = 1 - \alpha$.

The aggregation at the competency level was performed by using the classical OWA function. The associated weights were obtained through solving the same Equations (16)–(18) for $\alpha = 0.7$ and $n = 3$. Table 6 provides the optimal weighting vectors for different α values and $n = 3$. Note that values in Table 6 may not add up to one due to the conventions of rounding; however, the calculations were performed using original figures. Accordingly, the optimal three-dimensional OWA weighting vector was determined as $\mathbf{w} = (0.553955, 0.291992, 0.153999)$.

Table 6. The optimal weighting vector for $n = 3$ [72].

α	w_1	w_2	w_3
0.5	0.333333	0.333333	0.333333
0.6	0.438355	0.323242	0.238392
0.7	0.553955	0.291992	0.153999
0.8	0.681854	0.235840	0.081892
0.9	0.826294	0.146973	0.026306
1	1	0	0

Step 5.2. Aggregating CBDs for each candidate at question and competency level: First, for each candidate, the CBDs obtained for the interview questions associated with a particular competency were aggregated using Equations (21) and (22). The proposed aggregation operator requires, besides the optimal OWA weights, the accuracies of the classification models used. The idea is to assign higher weights to more accurate models. The accuracies were evaluated by the F1 score ranging from 0.74 to 0.84, as shown in Table 3. F1 score, which is calculated as the harmonic mean of the precision and recall of a test, performs well when there are imbalanced classes [73], as was the case in our example.

For instance, given the CBDs (γ_{11q4}) associated with the responses of Candidate 1 to the three interview questions of the competency “Communication”, the F1 scores (a_{1q}) of each classification model, and the OWA weighting vector (w_{1r}), the aggregation operation for the linguistic term s_4 is performed as follows:

$$\gamma_{114} = \left(\frac{1}{A} \sum_{q=1}^3 a_{1q} \gamma_{11q4} \sum_{r=1}^2 w_{1r} \gamma_{11(r)4} \right)^{\frac{1}{2}}$$

$$\gamma_{114} = \left[\frac{1}{2.34} (0.82 \cdot 0.32 \cdot (0.7 \cdot 0.64 + 0.3 \cdot 0) + 0.74 \cdot 0.64 \cdot (0.7 \cdot 0.32 + 0.3 \cdot 0) + 0.78 \cdot 0 \cdot (0.7 \cdot 0.64 + 0.3 \cdot 0.32)) \right]^{0.5} = 0.3091$$

The calculations for the remaining linguistic terms were made in a similar way and produced the following results $\gamma_{111} = 1$, $\gamma_{112} = 1$, $\gamma_{113} = 0.9675$, $\gamma_{115} = 0$. Table 7 provides the results for Candidate 1. The candidate has no low scores and achieved his best interview performance for the competency “Results Orientation” by exceeding expectations.

Table 7. Aggregated CBDs of Candidate 1.

Competency	s_1	s_2	s_3	s_4	s_5
Communication	1	1	0.9675	0.3091	0
Persuasiveness	1	1	0.4631	0	0
Results Orientation	1	1	1	0.8135	0.3083

The resulting aggregated cumulative belief structures were then aggregated at competency level to determine each candidate’s overall suitability for the open position. To do this, γ_{ijk} values were first ranked in descending order at each s_k level separately. Then, using Equation (23) with the optimal weighting vector $\mathbf{w} = (0.553955, 0.291992, 0.153999)$ obtained in Step 5.1, the final cumulative belief structures were calculated. For instance, the overall suitability of Candidate 1 at level s_4 was calculated as follows:

$$\gamma_{14} = \sum_{r=1}^3 w_r \gamma_{1(r)4} = 0.553955 \cdot 0.8135 + 0.291992 \cdot 0.3091 + 0.153999 \cdot 0 = 0.5409$$

With similar calculations for the remaining linguistic terms, the general CBD structure representing the overall suitability of Candidate 1 is $C_1 = \{(1.0000, s_1), (1.0000, s_2), (0.9078, s_3), (0.5409, s_4), (0.1708, s_5)\}$. Table 8 provides the results for all Candidates.

Table 8. Overall suitability of candidates.

Candidate	Final Cumulative Belief Structure					OSS *	LSS			
	s_1	s_2	s_3	s_4	s_5		$\tau=0.3$	$\tau=0.5$	$\tau=0.7$	$\tau=0.9$
4	1.0000	1.0000	1.0000	0.9496	0.4374	0.8774	s_5	s_4	s_4	s_4
6	1.0000	1.0000	1.0000	0.9813	0.3909	0.8744	s_5	s_4	s_4	s_4
10	1.0000	1.0000	0.9917	0.7679	0.2974	0.8114	s_4	s_4	s_4	s_3
2	1.0000	1.0000	0.9680	0.6410	0.0264	0.7270	s_4	s_4	s_3	s_3
1	1.0000	1.0000	0.9078	0.5409	0.1708	0.7239	s_4	s_4	s_3	s_3
9	1.0000	1.0000	0.9862	0.5749	0.0000	0.7122	s_4	s_4	s_3	s_3
5	1.0000	1.0000	0.9041	0.5651	0.0000	0.6938	s_4	s_4	s_3	s_3
11	1.0000	1.0000	0.6723	0.5358	0.0337	0.6483	s_4	s_4	s_2	s_2
12	1.0000	1.0000	0.9812	0.2595	0.0000	0.6481	s_3	s_3	s_3	s_3
8	1.0000	1.0000	0.6601	0.0000	0.0000	0.5320	s_3	s_3	s_2	s_2
3	1.0000	0.8867	0.1368	0.0000	0.0000	0.4047	s_2	s_2	s_2	s_1
7	1.0000	0.6700	0.0000	0.0000	0.0000	0.3340	s_2	s_2	s_1	s_1

*: Candidates are ranked according to their OSSs.

Step 6. Ranking Candidates: The final cumulative belief structures of the candidates were compared according to two approaches. First, the candidates were ranked based on their OSSs (see Table 8) obtained by converting the final cumulative belief structures into single values (see Equation (24)). For example, the OSS of Candidate 1 was calculated as follows:

$$\begin{aligned} \text{OSS}_1 &= \left[\left(\sum_{k=1}^4 v_k (\gamma_{1k} - \gamma_{1k+1}) \right) + v_5 \gamma_{15} \right] \frac{100}{5} \\ &= [1 \cdot (1 - 1) + 2 \cdot (1 - 0.9078) + 3 \cdot (0.9078 - 0.5409) + 4 \cdot (0.5409 - 0.1708) + 5 \cdot 0.1708] \cdot \frac{100}{5} = 0.7239 \end{aligned}$$

In the formula, the numerical values of each linguistic term were set to $v_k = k$, $k = 1, \dots, 5$. According to the OSSs, the candidates were ranked as $C_4 > C_6 > C_{10} > C_2 > C_1 > C_9 > C_5 > C_{11} > C_{12} > C_8 > C_3 > C_7$. Since a maximum of three candidates were planned to advance to the next stage of the interview process, the candidates C_4 , C_6 , and C_{10} with the highest OSSs, respectively, seemed to be promising. Additionally, the sudden change in OSSs down the rank order right after the first three candidates supported this conclusion (see Table 8).

For the Linguistic-Cut Approach, the final cumulative belief structures of the candidates were transformed into single linguistic terms using Equation (25). Four alternative thresholds (0.3, 0.5, 0.7, and 0.9) were considered to observe possible changes in the fulfillment of linguistic terms. For instance, the Linguistic Suitability Score of Candidate 1 for the threshold 0.7 was determined as follows:

$$\text{LSS}_1^{0.7} = \text{Sup}_{k=1, \dots, 5} [s_k | \gamma_{1k} \geq 0.7] = \text{Sup}_{k=1, \dots, 5} [s_1, s_2, s_3] = s_3$$

In this example, s_3 was the highest linguistic term with a CBD greater than or equal to 0.7. The complete distribution of the sufficiently fulfilled highest linguistic terms across different threshold values and candidates is given in Table 8. If the threshold was set to 0.3, eight candidates achieved LSSs that exceeded expectations or were exceptional. Similarly, if the threshold was set to 0.5, eight candidates exceeded expectations. However, if the threshold was set to 0.7, the number of candidates decreased to three. For higher thresholds, even fewer candidates remained who exceed expectations. The hiring team set s_4 as the minimum expectation for the open position to be met by the candidates. In addition, as mentioned before, a maximum of three candidates was planned to move forward to the next stage. Under these constraints, the most appropriate threshold was 0.7 indicating C_4 , C_6 , and C_{10} as the most promising candidates. Note that although it was also possible to determine a fixed threshold in advance, it would not allow this sensitivity check to be performed.

The graphical representation of the final cumulative belief structures, given in Figure 3, also indicates a three-candidate solution, which confirms the results obtained from the Aggregated Score Approach. In the figure, it can be noticed that after level s_3 the candidates 4, 6, and 10 are clearly dominating all other candidates.

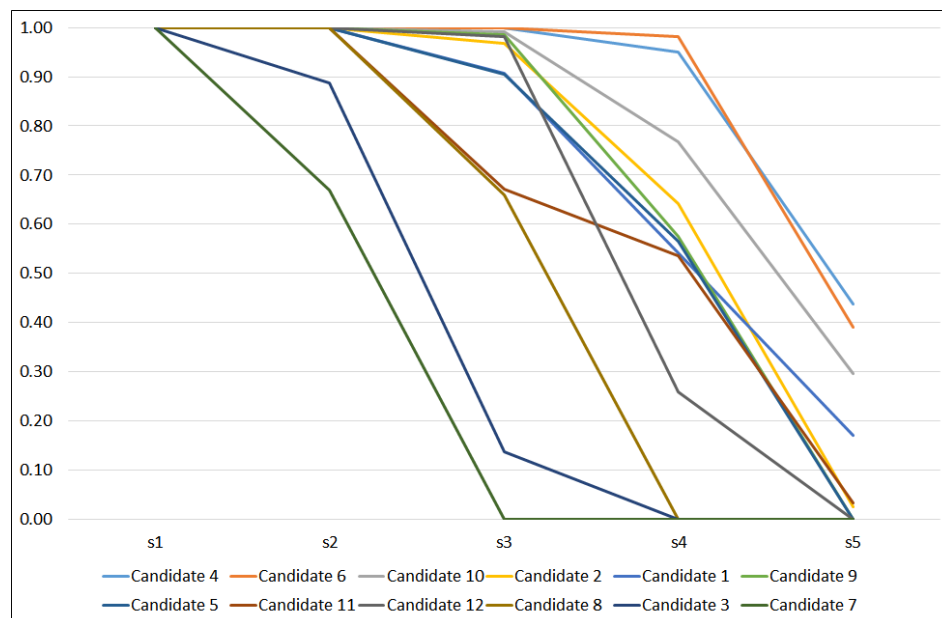


Figure 3. Comparison of the final CBDs of candidates.

Finally, to check the sensitivity of the ranking obtained by the proposed approach, an analysis of the rank stability in relation to the changes in weights was performed. Since the form of the weighting vector depends on the level of orness (α), the analysis was repeated with alternative orness degrees defined as $\alpha = 0$, $\alpha = 0.3$, $\alpha = 0.5$, $\alpha = 0.7$, and $\alpha = 1$. Here, the special cases are worth noting. For the orness degrees zero and one, the OWA function turns into the minimum and maximum operator, respectively. On the other hand, the level of orness set to 0.5 represents the special case where OWA is reduced to the usual average operator. Figure 4 shows the effect of variations of weights on the ranking of candidates. Comparison of the rankings indicates that although their rankings vary, candidates 4, 6, and 10 share the top three places in all cases. This can be simply explained by the high scores obtained by these candidates in almost all interview questions. The same robustness to the variations in weights can be observed on the other end of the ranking. Candidates 3 and 7 received relatively low scores in almost all interview questions and remained ignorant to changes in weights. However, for the remaining candidates (1, 2, 5, 8, 9, 11, and 12) the change in rankings becomes, as expected, more evident. This apparent change is caused by a few very low or very high scores obtained by these candidates from the interview questions. For example, Candidate 11 has mostly low scores and only a few were relatively high (refer to Table 3). While the low scores dominate the ranking at lower orness degrees, at higher levels of orness, the high scores start to be more influential (see Figure 4). Consequently, the obtained results and performed sensitivity analyses confirm the reliability and robustness of the proposed approach.

Step 7. Recommending Candidates: From the results of the assessment and selection process, it is clear that candidates 4, 6, and 10 deserved to advance to the next stage of the interview process. Examination of the candidates' predicted scores for each individual question (see Table 3) shows that Candidate 10 had strong communication skills, Candidate 6 had strong communication and persuasiveness skills, and Candidate 4 was strong in all skills. The hiring team reviewed the recommended candidates' interviews, resumes and suitability scores to confirm the shortlist of candidates. Consequently, candidates 4, 6, and 10 were invited for an online (synchronous) interview with the hiring manager.

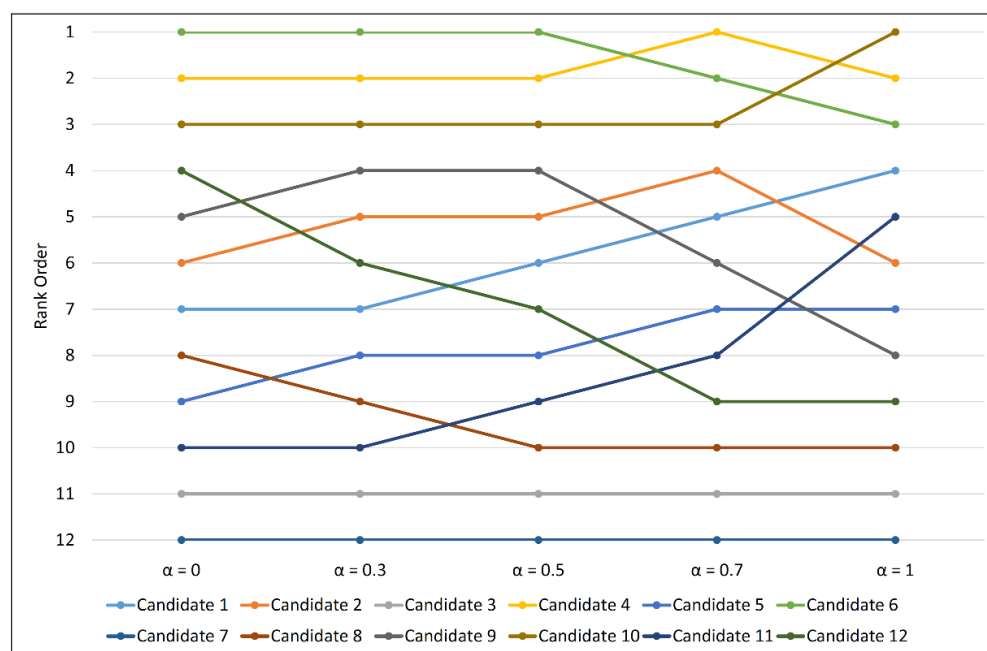


Figure 4. Sensitivity of ranking to variations in weights.

Comparative Analysis

To objectively evaluate the effectiveness of the proposed approach, the results were compared to the results of the rule-based CBD (RCBD) approach, classical Simple Additive Weighting (SAW), and Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS). The RCBD approach introduced by Kabak and Ruan [23] uses a rule-based system to aggregate CBDs. This approach, as described in Section 3, ignores incorporating interrelationships between criteria, considering accuracy of learning algorithms and weighting criteria objectively. For comparison purposes, the type of all interview questions were assumed important and, consistent with the proposed approach, the rule “two important questions are required to meet expectations for a competency at level s_i ” was defined. Finally, the max–min operator was used to aggregate the CBDs (for more detail, refer to [23]). According to SAW, a total score for each candidate was calculated by simply multiplying the (normalized) score for each response by the importance weight assigned to the question, followed by summing these products over all questions (for more details, the reader may refer to [74]). On the other hand, in TOPSIS [75] candidates are evaluated based on their distances to two reference points, namely the ideal best and the ideal worst. In this study, the Euclidean distances between the weighted normalized scores and the best and worst solutions achievable were used. The rationale for choosing the two classical MCDM techniques was that they are among the most cited and widely used methods in the personnel selection literature and are often used for comparison purposes (see [30,31,58,74], among others). Moreover, excluding the interrelationships and the OWA operator from the proposed approach will reduce the aggregation stage to a simple additive weighting technique. Thus, the comparison between the proposed approach and SAW will provide valuable evidence on the importance of considering interrelationships and using objective weighting. Another reason to restrict the comparisons to these two classical MCDM methods (SAW and TOPSIS) is that all other approaches presented in Section 2 require additional information such as pairwise comparison data, preference functions, and certain types of data (grey, intuitionistic, neutrosophic, etc.), which do not allow for direct comparisons.

To ensure a valid comparison between the results of the four approaches, the same assessment scores (predictions) and weights (accuracies), given in Table 3, were used. The results of the analyses in terms of OSSs and rankings are shown in Table 9. The comparisons show that while RCBD, SAW, and TOPSIS yielded highly similar rankings, the proposed

approach produced a rather different one. More precisely, in comparison to RCBD, SAW, and TOPSIS, the suggested approach delivered different rank orders for seven, seven, and eight candidates, respectively. The only candidates whose rank orders remained the same for all four approaches were candidates 3, 7, 8, and 10. This is not surprising; these candidates had the most stable rankings in the sensitivity analysis. The differences in rankings can be attributed to the (in)ability of the approaches to deal with uncertainty and subjectivity inherent in the assessments, as well as the interrelationships between the sub-criteria. Thus, the observed changes in the rankings provide evidence on the importance of using objective weighting and incorporating interrelationships into the analysis. Moreover, using a fuzzy linguistic term set to transform the scores also allowed consideration of the uncertainty inherent in the assessments.

Table 9. Comparison of the results.

Candidate	Proposed		RCBD		SWA		TOPSIS	
	OSS	Rank	OSS	Rank	OSS	Rank	OSS	Rank
1	0.7239	5	0.6667	7	0.6396	6	0.5522	7
2	0.7270	4	0.6907	4	0.6758	5	0.5939	5
3	0.4047	11	0.3787	11	0.2433	11	0.1534	11
4	0.8774	1	0.8480	2	0.8990	2	0.8356	2
5	0.6938	7	0.6693	6	0.6174	7	0.5533	6
6	0.8744	2	0.8720	1	0.9211	1	0.8556	1
7	0.3340	12	0.3253	12	0.1686	12	0.0529	12
8	0.5320	10	0.5120	10	0.4423	10	0.3552	10
9	0.7122	6	0.6827	5	0.6872	4	0.6357	4
10	0.8114	3	0.7627	3	0.7759	3	0.7078	3
11	0.6483	8	0.5600	9	0.5202	9	0.4424	9
12	0.6481	9	0.6480	8	0.6166	8	0.5475	8

Ranks in bold indicate pairwise disagreement between the proposed approach and the others.

Finally, to measure and statistically test the correspondence between the rankings of the four approaches, a correlation analysis was conducted. Since the ranks of the scores instead of the scores themselves are taken into consideration, a non-parametric measurement of relations should be used. Several measures are employed to evaluate the similarity between rankings, such as Kendall's rank correlation coefficient, Spearman's rank correlation coefficient, and Goodman and Kruskal's gamma coefficient. Recent studies have shown that Kendall's rank correlation coefficient is more robust, statistically more efficient, and mathematically more tractable (especially when ties are present) than Spearman's coefficient (for more details, see [76]). Therefore, in this study, Kendall's Tau-b statistic was preferred to determine how well the rankings of the techniques correlated with each other. It measures the monotonic relationship between rankings using the number of concordances and discordances in paired observations. It is calculated by the formula given below [77,78]:

$$\tau_b = \frac{N_c - N_d}{\sqrt{(N_0 - N_1)(N_0 - N_2)}} \quad (26)$$

$$N_0 = n(n-1)/2 \quad (27)$$

$$N_1 = \sum_i t_i(t_i - 1)/2 \quad (28)$$

$$N_2 = \sum_j u_j(u_j - 1)/2 \quad (29)$$

where N_c represents the number of concordant pairs, N_d represents the number of discordant pairs, t_i denotes the number of tied values in the i th group of ties for the first quantity, and u_j denotes the number of tied values in the j th group of ties for the second quantity. Kendall's tau-b ranges from -1 to 1 , where zero corresponds to no association and -1 and $+1$ both correspond to perfect association.

The results given in Table 10 show that all four approaches have statistically significant correlations with each other, yet the correlations between the proposed approach and the other three approaches are not as high as the correlations between RCBD and SAW ($\tau_b = 0.939$), RCBD and TOPSIS ($\tau_b = 0.970$), SAW and TOPSIS ($\tau_b = 0.970$), which are considerably high, as expected. The lowest correlation was observed between the proposed approach and TOPSIS ($\tau_b = 0.848$). The deviations from a perfect association with the classical approaches can be explained by the methodological differences introduced by the proposed approach, namely using objective weighting and incorporating interrelationships into the analysis.

Table 10. Correlation matrix (Kendall's tau b).

		Proposed	RCBD	SAW	TOPSIS
Proposed	Correlation Coefficient	1.000	0.879 **	0.879 **	0.848 **
	Significance (2-tailed)	.	0.000	0.000	0.000
	N	12	12	12	12
RCBD	Correlation Coefficient	0.879 **	1.000	0.939 **	0.970 **
	Significance (2-tailed)	0.000	.	0.000	0.000
	N	12	12	12	12
SAW	Correlation Coefficient	0.879 **	0.939 **	1.000	0.970 **
	Significance (2-tailed)	0.000	0.000	.	0.000
	N	12	12	12	12
TOPSIS	Correlation Coefficient	0.848 **	0.970 **	0.970 **	1.000
	Significance (2-tailed)	0.000	0.000	0.000	.
	N	12	12	12	12

** Correlation is significant at the 0.01 level (2-tailed).

Consequently, both the sensitivity and comparative analyses have demonstrated that using a CBD approach in combination with the weighted Bonferroni-OWA operator delivers satisfactory results. The findings in the application provide sufficient evidence for the effectiveness and applicability of the suggested approach.

The application presented here considered only verbal content to predict scores in the form of classification probabilities with respect to multiple interrelated questions using separate learning models. However, as previously mentioned, the proposed approach can be directly extended to handle AVI assessments in various forms obtained from both verbal and non-verbal content using different learning algorithms and models.

6. Conclusions and Further Research

AVI represents an emerging digital interview technology that more and more companies are adopting, especially in the early stages of the hiring process. In combination with AI and ML, AVI platforms provide the opportunity to assess video responses based on verbal and non-verbal indicators of the candidates to learn and predict patterns. Because of the complex, subjective, and uncertain nature of the personnel selection problem, many of the proposed learning-based approaches suffer from highly complex modeling and lack of large and labeled datasets. Moreover, models using sophisticated learning algorithms are often unable to explain the reasons for candidates' results. Therefore, considering all relevant (sub-)criteria in form of verbal and/or non-verbal cues in one single model or separately is a critical decision.

In order to address these issues, an effective, practical, and explainable approach is proposed that is able to transform, weight, combine, and rank automated AVI assessments obtained using multiple criteria. The proposed approach consists of two stages. The Assessments Stage learns and predicts patterns in candidate video responses using AVI in combination with learning-based approaches. The Selection Stage transforms and aggregates AVI assessment scores using an extended CBD approach in a multi-criteria environment. The proposed approach combines the benefits of learning-based techniques

with MCDM techniques and makes unique methodological contributions to the selection stage. The methodological contributions of the proposed approach are summarized below:

- This is the first study that combines wBON-OWA with the CBD approach to aggregate belief degrees in cumulative form. This integration allows us to capture the expressed interrelationship between the (sub-)criteria. Additionally, it incorporates the accuracy values of the learning algorithms as weights into the aggregation function to allow better models that have a greater effect on the final result. Using objective weighting of (sub-)criteria to give more importance to higher (or lower) scores is another significant contribution of this new approach.
- Using cumulative belief structures in combination with objective weighting and averaging ensures a completely data-driven and efficient approach to the whole selection process. In other words, depending on the preference of the hiring team, the process can be run without the need for any expert intervention and can efficiently deal with large numbers of (sub-)criteria and candidates.
- Depending on the ML algorithms used, the AVI assessment scores predicted for each response can be in different forms and scales (e.g., probability, distance, test score, linguistic term). The proposed approach transforms scores obtained in different forms into a common scale based on linguistic terms and thereby does not require any particular scaling properties of the data for aggregation.
- Since the proposed approach represents an aggregated score as a distribution over a predefined fuzzy linguistic term set, it provides further insight into the suitability of a candidate and allows consideration of the uncertainty inherent in the predictions.

In order to demonstrate the effectiveness and applicability of the proposed approach, an example case is presented in which candidates were selected in the early stages of the hiring process to be invited for an initial in-person interview. The results show considerable differences in the rankings of candidates between the proposed approach and classical MCDM approaches. The differences can be attributed to the proposed approach's ability to transform, weight, combine, and rank automated AVI assessments under multiple criteria.

The proposed approach also offers practical advantages to the hiring team. It enables the team to assess and select candidates more effectively and efficiently. The approach enables the team to automate and standardize many of the tasks that are typically required for candidate assessment and selection and minimize the subjectivity inherent in the process. This, in turn, increases the validity and fairness of the hiring process and increases the capacity for handling a large number of candidates. Using simpler separate prediction models for each interview question and combining the results using the proposed CBD approach provides not only a score/ranking but also the rationale behind it, which leads to more informed and better decisions.

There are also some limitations and potential future directions that should be acknowledged. The performance of the predictions in the assessment stage may also be influenced by AVI design features such as preparation and response length, option to re-record responses, and the method of presenting questions. However, the study remains limited to certain design features and left this issue as a concern for further research. Besides, ML models used in the assessment stage reduce some of the biases inherent in subjective assessments, but may suffer from biases arising from the development of these models. In particular, the quality and representativeness of the dataset from which to learn and the model used for training and prediction may influence the assessment of candidates and thus deserves further attention. Another future research direction could be examining the applicability and effectiveness of the proposed approach for the later stages of the selection process. As was the case in our example, automated assessments are often utilized in early stages of the assessment and selection process, but it is worth also considering later stages where more comprehensive assessments are required. It would be also interesting to apply the suggested approach to an augmented hiring process (see [12]) where AI supplements human decision-makers rather than replacing them. Since the proposed approach is able to flexibly transform and aggregate various forms of assessments, it can be easily extended to

this type of selection process without loss of generality. Furthermore, in order to transform assessments presented in different forms into belief structures, more advanced linguistic term sets (e.g., interval, hesitant) can be considered, which would be a challenging future study. Finally, developing an optimization model to determine the level of orness that distinguishes the candidates optimally given the hiring team's preferences (i.e., high or low orness degree) is also worth examining.

Author Contributions: Conceptualization, U.A. and A.S.; methodology, U.A. and A.S.; validation, U.A.; formal analysis, U.A.; investigation, U.A. and A.S.; writing—original draft preparation, U.A.; writing—review and editing, U.A. and A.S.; visualization, U.A.; project administration, U.A. and A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The Scientific and Technological Research Council of Turkey grant number 3150857.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors thank Şeyda Serdarasan for her helpful comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Belief degrees of Candidate 2.

Competency	Question	Prediction	Belief Degrees					CBDs				
			s_1	s_2	s_3	s_4	s_5	s_1	s_2	s_3	s_4	s_5
Communication ($j = 1$)	j_1	0.52			0.92	0.08		1	1	1	0.08	0
	j_2	0.62			0.52	0.48		1	1	1	0.48	0
	j_3	0.51			0.96	0.04		1	1	1	0.04	0
Persuasiveness ($j = 2$)	j_1	0.56			0.76	0.24		1	1	1	0.24	0
	j_2	0.65			0.40	0.60		1	1	1	0.60	0
	j_3	0.33		0.68	0.32			1	1	0.32	0	0
Results Orientation ($j = 3$)	j_1	0.78				0.88	0.12	1	1	1	1	0.12
	j_2	0.76				0.96	0.04	1	1	1	1	0.04
	j_3	0.73			0.08	0.92		1	1	1	0.92	0

Table A2. Belief degrees of Candidate 3.

Competency	Question	Prediction	Belief Degrees					CBDs				
			s_1	s_2	s_3	s_4	s_5	s_1	s_2	s_3	s_4	s_5
Communication ($j = 1$)	j_1	0.15	0.40	0.60				1	0.60	0	0	0
	j_2	0.22	0.12	0.88				1	0.88	0	0	0
	j_3	0.25		1				1	1	0	0	0
Persuasiveness ($j = 2$)	j_1	0.34		0.64	0.36			1	1	0.36	0	0
	j_2	0.28		0.88	0.12			1	1	0.12	0	0
	j_3	0.31		0.76	0.24			1	1	0.24	0	0
Results Orientation ($j = 3$)	j_1	0.10	0.60	0.40				1	0.40	0	0	0
	j_2	0.17	0.32	0.68				1	0.68	0	0	0
	j_3	0.14	0.44	0.56				1	0.56	0	0	0

Table A3. Belief degrees of Candidate 4.

Competency	Question	Prediction	Belief Degrees					CBDs				
			s_1	s_2	s_3	s_4	s_5	s_1	s_2	s_3	s_4	s_5
Communication ($j = 1$)	j_1	0.92				0.32	0.68	1	1	1	1	0.68
	j_2	0.87				0.52	0.48	1	1	1	1	0.48
	j_3	0.85				0.60	0.40	1	1	1	1	0.40
Persuasiveness ($j = 2$)	j_1	0.88				0.48	0.52	1	1	1	1	0.52
	j_2	0.90				0.40	0.60	1	1	1	1	0.60
	j_3	0.83				0.68	0.32	1	1	1	1	0.32
Results Orientation ($j = 3$)	j_1	0.71			0.16	0.84		1	1	1	0.84	0
	j_2	0.68			0.28	0.72		1	1	1	0.72	0
	j_3	0.60			0.60	0.40		1	1	1	0.40	0

Table A4. Belief degrees of Candidate 5.

Competency	Question	Prediction	Belief Degrees					CBDs				
			s_1	s_2	s_3	s_4	s_5	s_1	s_2	s_3	s_4	s_5
Communication ($j = 1$)	j_1	0.72			0.12	0.88		1	1	1	0.88	0
	j_2	0.70			0.20	0.80		1	1	1	0.80	0
	j_3	0.61			0.56	0.44		1	1	1	0.44	0
Persuasiveness ($j = 2$)	j_1	0.47		0.12	0.88			1	1	0.88	0	0
	j_2	0.73			0.08	0.92		1	1	1	0.92	0
	j_3	0.68			0.28	0.72		1	1	1	0.72	0
Results Orientation ($j = 3$)	j_1	0.38		0.48	0.52			1	1	0.52	0	0
	j_2	0.42		0.32	0.68			1	1	0.68	0	0
	j_3	0.27		0.92	0.08			1	1	0.08	0	0

Table A5. Belief degrees of Candidate 6.

Competency	Question	Prediction	Belief Degrees					CBDs				
			s_1	s_2	s_3	s_4	s_5	s_1	s_2	s_3	s_4	s_5
Communication ($j = 1$)	j_1	0.84				0.64	0.36	1	1	1	1	0.36
	j_2	0.91				0.36	0.64	1	1	1	1	0.64
	j_3	0.66			0.36	0.64		1	1	1	0.64	0
Persuasiveness ($j = 2$)	j_1	0.80				0.80	0.20	1	1	1	1	0.20
	j_2	0.74			0.04	0.96		1	1	1	0.96	0
	j_3	0.85				0.60	0.40	1	1	1	1	0.40
Results Orientation ($j = 3$)	j_1	0.83				0.68	0.32	1	1	1	1	0.32
	j_2	0.88				0.48	0.52	1	1	1	1	0.52
	j_3	0.89				0.44	0.56	1	1	1	1	0.56

Table A6. Belief degrees of Candidate 7.

Competency	Question	Prediction	Belief Degrees					CBDs				
			s_1	s_2	s_3	s_4	s_5	s_1	s_2	s_3	s_4	s_5
Communication ($j = 1$)	j_1	0.18	0.28	0.72				1	0.72	0	0	0
	j_2	0.00	1					1	0	0	0	0
	j_3	0.14	0.44	0.56				1	0.56	0	0	0
Persuasiveness ($j = 2$)	j_1	0.23	0.08	0.92				1	0.92	0	0	0
	j_2	0.16	0.36	0.64				1	0.64	0	0	0
	j_3	0.10	0.60	0.40				1	0.40	0	0	0
Results Orientation ($j = 3$)	j_1	0.17	0.32	0.68				1	0.68	0	0	0
	j_2	0.13	0.48	0.52				1	0.52	0	0	0
	j_3	0.24	0.04	0.96				1	0.96	0	0	0

Table A7. Belief degrees of Candidate 8.

Competency	Question	Prediction	Belief Degrees					CBDs				
			s_1	s_2	s_3	s_4	s_5	s_1	s_2	s_3	s_4	s_5
Communication ($j = 1$)	j_1	0.42		0.32	0.68			1	1	0.68	0	0
	j_2	0.50			1			1	1	1	0	0
	j_3	0.39		0.44	0.56			1	1	0.56	0	0
Persuasiveness ($j = 2$)	j_1	0.36		0.56	0.44			1	1	0.44	0	0
	j_2	0.43		0.28	0.72			1	1	0.72	0	0
	j_3	0.38		0.48	0.52			1	1	0.52	0	0
Results Orientation ($j = 3$)	j_1	0.37		0.52	0.48			1	1	0.48	0	0
	j_2	0.42		0.32	0.68			1	1	0.68	0	0
	j_3	0.30		0.80	0.20			1	1	0.20	0	0

Table A8. Belief degrees of Candidate 9.

Competency	Question	Prediction	Belief Degrees					CBDs				
			s_1	s_2	s_3	s_4	s_5	s_1	s_2	s_3	s_4	s_5
Communication ($j = 1$)	j_1	0.62			0.52	0.48		1	1	1	0.48	0
	j_2	0.64			0.44	0.56		1	1	1	0.56	0
	j_3	0.75				1		1	1	1	1	0
Persuasiveness ($j = 2$)	j_1	0.56			0.76	0.24		1	1	1	0.24	0
	j_2	0.60			0.60	0.40		1	1	1	0.40	0
	j_3	0.73			0.08	0.92		1	1	1	0.92	0
Results Orientation ($j = 3$)	j_1	0.42		0.32	0.68			1	1	0.68	0	0
	j_2	0.57			0.72	0.28		1	1	1	0.28	0
	j_3	0.65			0.40	0.60		1	1	1	0.60	0

Table A9. Belief degrees of Candidate 10.

Competency	Question	Prediction	Belief Degrees					CBDs				
			s_1	s_2	s_3	s_4	s_5	s_1	s_2	s_3	s_4	s_5
Communication ($j = 1$)	j_1	0.89				0.44	0.56	1	1	1	1	0.56
	j_2	0.80				0.80	0.20	1	1	1	1	0.20
	j_3	0.95				0.20	0.80	1	1	1	1	0.80
Persuasiveness ($j = 2$)	j_1	0.68			0.28	0.72		1	1	1	0.72	0
	j_2	0.45		0.20	0.80			1	1	0.80	0	0
	j_3	0.59			0.64	0.36		1	1	1	0.36	0
Results Orientation ($j = 3$)	j_1	0.63			0.48	0.52		1	1	1	0.52	0
	j_2	0.70			0.20	0.80		1	1	1	0.80	0
	j_3	0.57			0.72	0.28		1	1	1	0.28	0

Table A10. Belief degrees of Candidate 11.

Competency	Question	Prediction	Belief Degrees					CBDs				
			s_1	s_2	s_3	s_4	s_5	s_1	s_2	s_3	s_4	s_5
Communication ($j = 1$)	j_1	0.26		0.96	0.04			1	1	0.04	0	0
	j_2	0.35		0.60	0.40			1	1	0.40	0	0
	j_3	0.25		1				1	1	0	0	0
Persuasiveness ($j = 2$)	j_1	0.43		0.28	0.72			1	1	0.72	0	0
	j_2	0.33		0.68	0.32			1	1	0.32	0	0
	j_3	0.27		0.92	0.08			1	1	0.08	0	0
Results Orientation ($j = 3$)	j_1	0.76				0.96	0.04	1	1	1	1	0.04
	j_2	0.72			0.12	0.88		1	1	1	0.88	
	j_3	0.80				0.80	0.20	1	1	1	1	0.20

Table A11. Belief degrees of Candidate 12.

Competency	Question	Prediction	Belief Degrees					CBDs				
			s ₁	s ₂	s ₃	s ₄	s ₅	s ₁	s ₂	s ₃	s ₄	s ₅
Communication (j = 1)	j ₁	0.48		0.08	0.92			1	1	0.92	0	0
	j ₂	0.63			0.48	0.52		1	1	1	0.52	0
	j ₃	0.55			0.80	0.20		1	1	1	0.20	0
Persuasiveness (j = 2)	j ₁	0.56			0.76	0.24		1	1	1	0.24	0
	j ₂	0.63			0.48	0.52		1	1	1	0.52	0
	j ₃	0.43		0.28	0.72			1	1	0.72	0	0
Results Orientation (j = 3)	j ₁	0.52			0.92	0.08		1	1	1	0.08	0
	j ₂	0.62			0.52	0.48		1	1	1	0.48	0
	j ₃	0.57			0.72	0.28		1	1	1	0.28	0

References

- Slama, M.Y. *Human Resources Professionals' Perceptions and Use of Asynchronous Video Interviewing*; Saint Mary's University: Halifax, NS, Canada, 2020.
- Deloitte, I. Leading the Social Enterprise: Reinvent with a Human Focus: 2019 Deloitte Global Human Capital Trends. 2019. Available online: https://www2.deloitte.com/content/dam/insights/us/articles/5136_HC-Trends-2019/DI_HC-Trends-2019.pdf (accessed on 20 October 2021).
- Kiron, D.; Prentice, P.K.; Ferguson, R.B. Raising the bar with analytics. *MIT Sloan Manag. Rev.* **2014**, *55*, 29.
- Nikolaou, I. What is the Role of Technology in Recruitment and Selection? *Span. J. Psychol.* **2021**, *24*, 1–6. [CrossRef] [PubMed]
- Smith, A.D.; Rupp, W.T. Managerial challenges of e-recruiting: Extending the life cycle of new economy employees. *Online Inf. Rev.* **2004**, *28*, 61–74. [CrossRef]
- Chen, C.; Lü, J.; Shen, H. Fine-Grained Interview Evaluation Method Based on Keyword Attention. *J. Comput. Res. Dev.* **2021**, *58*, 2013–2024.
- Mejia, C.; Torres, E.N. Implementation and normalization process of asynchronous video interviewing practices in the hospitality industry. *Int. J. Contemp. Hosp. Manag.* **2018**, *30*, 685–701. [CrossRef]
- Rao, S.B.P.; Rasipuram, S.; Das, R.; Jayagopi, D.B. Automatic assessment of communication skill in non-conventional interview settings: A comparative study. In Proceedings of the 19th ACM International Conference on Multimodal Interaction, Glasgow, UK, 13–17 November 2017; pp. 221–229.
- Rubinstein, P. Asynchronous Video Interviews: The Tools You Need to Succeed—BBC Worklife. 2020. Available online: <https://www.bbc.com/worklife/article/20201102-asynchronous-video-interviews-the-tools-you-need-to-succeed> (accessed on 20 October 2021).
- Maurer, R. Digital Video Upgrades the Hiring Experience. 2017. Available online: <https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/digital-video-upgrades-the-hiring-experience.aspx> (accessed on 3 September 2021).
- Lukacik, E.-R.; Bourdage, J.S.; Roulin, N. Into the void: A conceptual model and research agenda for the design and use of asynchronous video interviews. *Hum. Resour. Manag. Rev.* **2022**, *32*, 100789. [CrossRef]
- Gonzalez, M.F.; Liu, W.; Shirase, L.; Tomczak, D.L.; Lobbe, C.E.; Justenhoven, R.; Martin, N.R. Allying with AI? Reactions toward human-based, AI/ML-based, and augmented hiring processes. *Comput. Hum. Behav.* **2022**, *130*, 107179. [CrossRef]
- Chen, L.; Feng, G.; Leong, C.W.; Lehman, B.; Martin-Raugh, M.; Kell, H.; Lee, C.M.; Yoon, S.-Y. Automated scoring of interview videos using Doc2Vec multimodal feature extraction paradigm. In Proceedings of the 18th ACM International Conference on Multimodal Interaction, Tokyo, Japan, 12–16 November 2016; pp. 161–168.
- HireVue. Available online: www.hirevue.com (accessed on 10 April 2022).
- Modern Hire. Available online: modernhire.com (accessed on 10 April 2022).
- Talview. Available online: www.talview.com (accessed on 10 April 2022).
- Black, J.S.; van Esch, P. AI-enabled recruiting: What is it and how should a manager use it? *Bus. Horiz.* **2020**, *63*, 215–226. [CrossRef]
- Kim, J.-Y.; Heo, W. Artificial intelligence video interviewing for employment: Perspectives from applicants, companies, developer and academicians. *Inf. Technol. People* **2021**, *35*, 861–878. [CrossRef]
- Garg, S.; Sinha, S.; Kar, A.K.; Mani, M. A review of machine learning applications in human resource management. *Int. J. Product. Perform. Manag.* **2021**, *71*, 1590–1610. [CrossRef]
- Rasipuram, S.; Jayagopi, D.B. Automatic multimodal assessment of soft skills in social interactions: A review. *Multimed. Tools Appl.* **2020**, *79*, 13037–13060. [CrossRef]
- Baltrušaitis, T.; Ahuja, C.; Morency, L.-P. Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 423–443. [CrossRef] [PubMed]
- Elgendy, N.; Elragal, A.; Päiväranta, T. DECAS: A modern data-driven decision theory for big data and analytics. *J. Decis. Syst.* **2021**, 1–37. [CrossRef]

23. Kabak, Ö.; Ruan, D.A. A cumulative belief degree-based approach for missing values in nuclear safeguards evaluation. *IEEE Trans. Knowl. Data Eng.* **2011**, *23*, 1441–1454. [\[CrossRef\]](#)
24. Yager, R.R. On generalized Bonferroni mean operators for multi-criteria aggregation. *Int. J. Approx. Reason.* **2009**, *50*, 1279–1286. [\[CrossRef\]](#)
25. Hemamou, L.; Felhi, G.; Vandenbussche, V.; Martin, J.-C.; Clavel, C. Hirenet: A hierarchical attention model for the automatic analysis of asynchronous video job interviews. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 573–581.
26. Chen, L.; Feng, G.; Martin-Raugh, M.P.; Leong, C.W.; Kitchen, C.; Yoon, S.Y.; Lehman, B.; Kell, H.; Lee, C.M. Automatic Scoring of Monologue Video Interviews Using Multimodal Cues. In Proceedings of the INTERSPEECH, San Francisco, CA, USA, 8–12 September 2016; pp. 32–36.
27. Rasipuram, S.; Rao, P.S.B.; Jayagopi, D.B. Asynchronous video interviews vs. face-to-face interviews for communication skill measurement: A systematic study. In Proceedings of the 18th ACM International Conference on Multimodal Interaction, Tokyo, Japan, 12–16 November 2016; pp. 370–377.
28. Chen, L.; Zhao, R.; Leong, C.W.; Lehman, B.; Feng, G.; Hoque, M.E. Automated video interview judgment on a large-sized corpus collected online. In Proceedings of the 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), San Antonio, TX, USA, 23–26 October 2017; pp. 504–509.
29. Chen, K.; Niu, M.; Chen, Q. A Hierarchical Reasoning Graph Neural Network for The Automatic Scoring of Answer Transcriptions in Video Job Interviews. *Int. J. Mach. Learn. Cybern.* **2022**, 1–11. [\[CrossRef\]](#)
30. Chen, C.-T.; Hung, W.-Z. A two-phase model for personnel selection based on multi-type fuzzy information. *Mathematics* **2020**, *8*, 1703. [\[CrossRef\]](#)
31. Luo, S.; Xing, L. A hybrid decision making framework for personnel selection using BWM, MABAC and PROMETHEE. *Int. J. Fuzzy Syst.* **2019**, *21*, 2421–2434. [\[CrossRef\]](#)
32. Chuang, Y.-C.; Hu, S.-K.; Liou, J.J.H.; Tzeng, G.-H. A data-driven MADM model for personnel selection and improvement. *Technol. Econ. Dev. Econ.* **2020**, *26*, 751–784. [\[CrossRef\]](#)
33. Kilic, H.S.; Demirci, A.E.; Delen, D. An integrated decision analysis methodology based on IF-DEMATEL and IF-ELECTRE for personnel selection. *Decis. Support Syst.* **2020**, *137*, 113360. [\[CrossRef\]](#)
34. Yalçın, N.; Pehlivan, N.Y. Application of the fuzzy CODAS method based on fuzzy envelopes for hesitant fuzzy linguistic term sets: A case study on a personnel selection problem. *Symmetry* **2019**, *11*, 493. [\[CrossRef\]](#)
35. Karabašević, D.; Stanujkić, D.; Urošević, S. The MCDM Model for Personnel Selection Based on SWARA and ARAS Methods. *Management* **2015**, *20*, 43–52.
36. Karabasevic, D.; Zavadskas, E.K.; Stanujkic, D.; Popovic, G.; Brzakovic, M. An Approach to Personnel Selection in the IT Industry Based on the EDAS Method. *Transform. Bus. Econ.* **2018**, *17*, 54–65.
37. Liu, H.-C.; Qin, J.-T.; Mao, L.-X.; Zhang, Z.-Y. Personnel Selection Using Interval 2-Tuple Linguistic VIKOR Method. *Hum. Factors Ergon. Manuf. Serv. Ind.* **2015**, *25*, 370–384. [\[CrossRef\]](#)
38. Sang, X.; Liu, X.; Qin, J. An analytical solution to fuzzy TOPSIS and its application in personnel selection for knowledge-intensive enterprise. *Appl. Soft Comput.* **2015**, *30*, 190–204. [\[CrossRef\]](#)
39. Ji, P.; Zhang, H.; Wang, J. A projection-based TODIM method under multi-valued neutrosophic environments and its application in personnel selection. *Neural Comput. Appl.* **2018**, *29*, 221–234. [\[CrossRef\]](#)
40. Özgörmüş, E.; Şenocak, A.A.; Gören, H.G. An integrated fuzzy QFD-MCDM framework for personnel selection problem. *Sci. Iran.* **2021**, *28*, 2972–2986.
41. Krishankumar, R.; Premaladha, J.; Ravichandran, K.S.; Sekar, K.R.; Manikandan, R.; Gao, X.Z. A novel extension to VIKOR method under intuitionistic fuzzy context for solving personnel selection problem. *Soft Comput.* **2020**, *24*, 1063–1081. [\[CrossRef\]](#)
42. Ulutaş, A.; Popovic, G.; Stanujkic, D.; Karabasevic, D.; Zavadskas, E.K.; Turskis, Z. A new hybrid MCDM model for personnel selection based on a novel grey PIPRECIA and grey OCRA methods. *Mathematics* **2020**, *8*, 1698. [\[CrossRef\]](#)
43. Herrera, F.; Herrera-Viedma, E.; Martínez, L. A fusion approach for managing multi-granularity linguistic term sets in decision making. *Fuzzy Sets Syst.* **2000**, *114*, 43–58. [\[CrossRef\]](#)
44. Kabak, Ö.; Ruan, D.A. A comparison study of fuzzy MADM methods in nuclear safeguards evaluation. *J. Glob. Optim.* **2011**, *51*, 209–226. [\[CrossRef\]](#)
45. Ervural, B.; Kabak, Ö. A cumulative belief degree approach for group decision-making problems with heterogeneous information. *Expert Syst.* **2019**, *36*, e12458. [\[CrossRef\]](#)
46. Ünüçay, H.; Ervural, B.Ç.; Ervural, B.; Kabak, Ö. Cumulative belief degrees approach for assessment of sustainable development. In *Intelligence Systems in Environmental Management: Theory and Applications*; Kahraman, C., Sari, I.U., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 257–289.
47. Gül, S.; Kabak, Ö.; Topcu, I. A multiple criteria credit rating approach utilizing social media data. *Data Knowl. Eng.* **2018**, *116*, 80–99. [\[CrossRef\]](#)
48. Ruan, D.; Kabak, Ö.; Quinones, R. An ordered weighted averaging operator-based cumulative belief degree approach for energy policy evaluation. *Int. J. Adv. Oper. Manag.* **2013**, *5*, 58–73. [\[CrossRef\]](#)
49. Zorluoğlu, Ö.Ş.; Kabak, Ö. Weighted cumulative belief degree approach for project portfolio selection. *Gr. Decis. Negot.* **2020**, *29*, 679–722. [\[CrossRef\]](#)

50. Kabak, Ö.; Cinar, D.; Hoge, G.Y. A Cumulative Belief Degree Approach for Prioritization of Energy Sources: Case of Turkey. In *Assessment and Simulation Tools for Sustainable Energy Systems: Theory and Applications*; Cavallaro, F., Ed.; Springer: London, UK, 2013; pp. 129–151.
51. Bozdag, E.; Asan, U.; Soyer, A.; Serdarasan, S. Risk prioritization in failure mode and effects analysis using interval type-2 fuzzy sets. *Expert Syst. Appl.* **2015**, *42*, 4000–4015. [\[CrossRef\]](#)
52. Gül, S.; Kabak, Ö.; Topcu, Y.I. An OWA operator-based cumulative belief degrees approach for credit rating. *Int. J. Intell. Syst.* **2018**, *33*, 998–1026. [\[CrossRef\]](#)
53. Yager, R.R. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Trans. Syst. Man. Cybern.* **1988**, *18*, 183–190. [\[CrossRef\]](#)
54. Beliakov, G.; James, S.; Mesiar, R. A generalization of the Bonferroni mean based on partitions. In Proceedings of the 2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Hyderabad, India, 7–10 July 2013; pp. 1–6.
55. Bonferroni, C. Sulle medie multiple di potenze. *Boll. Dell'unione Mat. Ital.* **1950**, *5*, 267–270.
56. Perez-Arellano, L.A.; Blanco-Mesa, F.; Leon-Castro, E.; Alfaro-Garcia, V. Bonferroni prioritized aggregation operators applied to government transparency. *Mathematics* **2021**, *9*, 24. [\[CrossRef\]](#)
57. Espinoza-Audelo, L.F.; Olazabal-Lugo, M.; Blanco-Mesa, F.; León-Castro, E.; Alfaro-Garcia, V. Bonferroni probabilistic ordered weighted averaging operators applied to agricultural commodities' price analysis. *Mathematics* **2020**, *8*, 1350. [\[CrossRef\]](#)
58. Chen, Z.-S.; Zhang, X.; Rodríguez, R.M.; Wang, X.; Chin, K.-S. Heterogeneous Interrelationships among Attributes in Multi-Attribute Decision-Making: An Empirical Analysis. *Int. J. Comput. Intell. Syst.* **2019**, *12*, 984–997. [\[CrossRef\]](#)
59. Beliakov, G.; James, S.; Mordelova, J.; Rueckschlossova, T.; Yager, R.R. Generalized Bonferroni mean operators in multi-criteria aggregation. *Fuzzy Sets Syst.* **2010**, *161*, 2227–2242. [\[CrossRef\]](#)
60. O'Hagan, M. Aggregating template or rule antecedents in real-time expert systems with fuzzy set logic. In Proceedings of the 22nd Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 31 October–2 November 1988; Volume 2, pp. 681–689.
61. Fuller, R.; Majlender, P. An analytic approach for obtaining maximal entropy OWA operator weights. *Fuzzy Sets Syst.* **2001**, *124*, 53–57. [\[CrossRef\]](#)
62. Filev, D.; Yager, R.R. On the issue of obtaining OWA operator weights. *Fuzzy Sets Syst.* **1998**, *94*, 157–169. [\[CrossRef\]](#)
63. Xu, Z.S.; Da, Q.-L. The uncertain OWA operator. *Int. J. Intell. Syst.* **2002**, *17*, 569–575. [\[CrossRef\]](#)
64. Xu, Z. An overview of methods for determining OWA weights. *Int. J. Intell. Syst.* **2005**, *20*, 843–865. [\[CrossRef\]](#)
65. Lin, M.; Xu, W.; Lin, Z.; Chen, R. Determine OWA operator weights using kernel density estimation. *Econ. Res. Istraživanja* **2020**, *33*, 1441–1464. [\[CrossRef\]](#)
66. Dominguez-Catena, I.; Paternain, D.; Galar, M. A Study of OWA Operators Learned in Convolutional Neural Networks. *Appl. Sci.* **2021**, *11*, 7195. [\[CrossRef\]](#)
67. Yager, R.R. On the dispersion measure of OWA operators. *Inf. Sci.* **2009**, *179*, 3908–3919. [\[CrossRef\]](#)
68. Kaufmann, A.; Gupta, M.M. *Introduction to Fuzzy Arithmetic: Theory and Applications*; Van Nostrand Reinhold: New York, NY, USA, 1991.
69. Google Speech-to-Text API. Available online: <https://cloud.google.com/speech-to-text/> (accessed on 10 April 2022).
70. Jurafsky, D.; Martin, J.H. *Speech and Language Processing* (3rd ed. Draft). 2021. Available online: <https://web.stanford.edu/~jurafsky/slp3> (accessed on 10 February 2022).
71. fastText. Available online: <https://fasttext.cc/docs/en/support.html> (accessed on 10 April 2022).
72. Chang, K.-H.; Wen, T.-C. A novel efficient approach for DFMEA combining 2-tuple and the OWA operator. *Expert Syst. Appl.* **2010**, *37*, 2362–2370. [\[CrossRef\]](#)
73. Witten, I.H.; Frank, E.; Hall, M.A. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, 3rd ed.; Morgan Kaufmann Pub: Burlington, MA, USA, 2011.
74. Afshari, A.; Mojahed, M.; Yusuff, R.M. Simple additive weighting approach to personnel selection problem. *Int. J. Innov. Manag. Technol.* **2010**, *1*, 511.
75. Hwang, C.-L.; Yoon, K. Methods for Multiple Attribute Decision Making. In *Multiple Attribute Decision Making. Lecture Notes in Economics and Mathematical Systems*; Hwang, C.-L., Yoon, K., Eds.; Springer: Berlin/Heidelberg, Germany, 1981; pp. 58–191.
76. Croux, C.; Dehon, C. Influence functions of the Spearman and Kendall correlation measures. *Stat. Methods Appl.* **2010**, *19*, 497–515. [\[CrossRef\]](#)
77. Kendall, M. *Rank Correlation Methods*, 4th ed.; Charles Griffin: London, UK, 1970.
78. Zamani-Sabzi, H.; King, J.P.; Gard, C.C.; Abudu, S. Statistical and analytical comparison of multi-criteria decision-making techniques under fuzzy environment. *Oper. Res. Perspect.* **2016**, *3*, 92–117. [\[CrossRef\]](#)