A Distance-metric Approach to Expert Agreement In Multi-level Publications Classification*

Ihor Turkin[1],†, Andriy Chukhray[1],*,†, Oleksandr Liubimov[1],† and Lina Volobuieva[1],†

[1]*National Aerospace University "Kharkiv Aviation Institute", 17 Vadym Manko St, Kharkiv, 61070, Ukraine*

Abstract

In secondary scientific research, there is often a need to classify publications by content. However, the results of different experts may differ significantly due to different experiences, context, or individual biases. This creates difficulties in the formation of systematic reviews and scientific analytics. The article proposes a method for agreeing expert assessments that takes into account the multi-level classification structure (main and additional classes) and the relative weight of the selected categories. To quantify the differences, a distance metric was used, built by analogy with the Levenshtein distance, with the determination of the cost of editing operations (Replace, Rotate, Add). The proposed approach allows finding a compromise agreed result even with significant differences in assessments. The method was tested on a case of classification of 203 scientific publications on the architecture of on-board computers of CubeSat nanosatellites. The results showed that the use of the method reduces the number of conflict cases and increases the stability of the classification compared to simple majority voting. The proposed approach can be used not only in scientometrics, but also in expert systems and group decision-making tasks.

Keywords

expert classification, consensus evaluation, distance metric, systematic review, artificial intelligence

1. Introduction

In the modern scientific environment, the number of publications is growing exponentially. This applies to almost all fields of knowledge - from biomedicine to aerospace technologies. Only in the field of development of on-board computers for CubeSat nanosatellites, hundreds of scientific papers are registered every year in leading databases such as Scopus, IEEE Xplore, ACM Digital Library, Web of Science, etc. There is an objective need to systematize the knowledge obtained and form generalized conclusions based on heterogeneous research. One of the common approaches to such systematization is Systematic Literature Review (SLR). It is based on a clear methodology for selecting, analyzing, and summarizing scientific publications. Expert classification of publications plays an important role in this process: specialists in the relevant subject area analyze abstracts or full texts of works and assign them to certain classes (topics, areas, subfields). However, the key problem with this approach is the discrepancies between expert assessments. The reasons for these differences may be different:

- different levels of knowledge and experience of experts;
- emphasis on specific aspects of research (for example, hardware or software);
- subjective biases or interests;
- the multi-topic and interdisciplinary nature of the publications themselves.

It leads to a situation where the same article can be assigned to different classes by different experts, or when the priority of classes is determined differently. This reduces the objectivity of the results and also makes it difficult to draw reliable scientometric conclusions.

ProfIT AI'25: 5th International Workshop of IT-professionals on Artificial Intelligence, October 15–17, 2025, Liverpool, UK

*

*Corresponding author

†These authors contributed equally.

i.turkin@khai.edu (I. Turkin); a.chukhray@khai.edu (A. Chukhray); oleksandr.liubimov@gmail.com (O. Liubimov); l.volobuieva@khai.edu (L. Volobuieva)

0000-0002-3986-4186 (I. Turkin); 0000-0002-8075-3664 (A. Chukhray); 0000-0003-3636-6939 (O. Liubimov); 0000-0002-3466-5743 (L. Volobuieva)

Traditionally, methods such as majority voting or calculation of agreement coefficients (e.g., Cohen's Kappa, Fleiss' Kappa [1]) have been used to reconcile expert opinions. However, these approaches have significant limitations, as:

1. Do not take into account the multi-level nature of expert assessments (main and additional classes).
2. Lose information about the hierarchy of importance.
3. Often reduces a complex decision-making process to a simple choice of one category.

Thus, there is a need for a method of coordinating expert assessments that combines formalization in the form of a mathematical model and preserves the informativeness of classifications. In this paper, we propose a method based on a distance metric between expert assessments, similar to the Levenshtein distance. The idea is to determine the minimum cost of transforming one expert assessment into another, given the allowed operations:

- **Replace** (class replacement);
- **Rotate** (change of class order);
- **Add** (adding a new class).

This approach allows:

- quantify the differences between estimates;
- build a normalized metric of the quality of expert evaluation;
- find a balanced option that minimizes the distance to the estimates of all experts.

Unlike simple methods, our approach preserves the multi-layered nature of the classification and allows us to work with interdisciplinary publications, when several areas are considered simultaneously in one article. To test the method, a subject area related to the development of the architecture of CubeSat onboard computers was chosen [2]. This is a relevant direction, since modern satellite missions require innovative hardware and software solutions, and the number of studies in this area is rapidly growing. As a test dataset, 203 publications over the past ten years were selected from the IEEE Xplore and Scopus databases. Two independent experts classified these works according to the specified classes, after which the proposed method of agreeing on estimates was applied. The results show that the new method significantly reduces the number of conflicting cases and produces a more stable and objective classification. In particular, only 18 out of 203 articles required re-review after automated reconciliation. Overall, this work contributes to two areas:

1. **Methodological** — a new method of coordinating expert assessments has been proposed, which can be used not only in scientometrics, but also in peer review and expert systems.
2. **Applied** — the effectiveness of the method was demonstrated on a real case of classification of publications about CubeSat on-board computers, which confirms its practical value in interdisciplinary research.

Thus, the article lays the foundation for further development of methods for agreeing expert classifications in the context of artificial intelligence and machine learning, where the task of building consensus is key to increasing the accuracy and objectivity of decisions.

2. State-of-the-art

The problem of consensus among experts is a classic problem in the field of collective decision-making, bibliometrics, and artificial intelligence. The scientific literature describes a number of approaches that allow assessing the degree of consensus between experts and for§ming a collective decision.

Traditional statistical methods. The most common statistical measures of consistency are given below:

- **Kappa coefficient** (Cohen's Kappa, Fleiss' Kappa) is used to measure the level of agreement between two or more experts, taking into account chance coincidences. This approach allows us to determine how much better the experts' estimates are than chance guessing, but does not provide a mechanism for forming an integrated result.
- **Kendall's W** [3] is used to analyze rankings provided by experts. It shows the level of concordance between orders, but does not take into account the structure of the classes themselves and may be insensitive to multilevel assessments.
- Methods based on arithmetic mean or majority voting [4] allow for rapid integration of estimates, but often lead to loss of information, especially in cases where the publication is interdisciplinary in nature and may belong to several classes simultaneously.

Consensus-based and collective learning approaches. In the field of machine learning, the problem of consensus estimation is often considered as an ensemble learning problem [5]. The idea is to combine the results of several "classifiers" to improve the accuracy of the final solution. Some well-known methods include:

- bagging and boosting — use iterative learning and weight adjustment of classifier results;
- stacking — combines multiple models through a metaclassifier;
- Consensus clustering [6] is used in clustering problems, where the goal is to construct a consistent partition of a set of objects based on several independent clusterings.

These approaches show high efficiency, but their direct application to expert assessments has limitations: experts often operate not only with one "label", but with ordered sets of classes (main class, additional classes), which requires preserving the hierarchy.

Methods in bibliometrics and scientometrics. In bibliometric studies, there is also a need to develop agreed classifications. Some studies suggest using the Delphi method [7], where experts gradually revise their assessments until an acceptable level of agreement is reached. The advantage is that consensus can be gradually reached, but the disadvantage is that it is time-consuming. Another direction is related to the application of multi-criteria decision-making (MCDM) methods [8], such as AHP (Analytic Hierarchy Process), TOPSIS, or ELECTRE. They allow building agreed solutions based on weight coefficients and a multi-level structure of criteria, but require formalization of preferences of all experts, which is not always possible in publication classification problems.

Identified limitations. Analysis of the literature allows us to highlight several key problems:

1. Information loss: simple methods (majority voting) ignore secondary classes.
2. Low sensitivity to hierarchy: Consistency metrics (Kappa, Kendall's W) work with nominal or ranked data, but do not take into account ordered sets of classes.
3. High labor intensity: Delphi-like approaches are time-consuming and do not scale for large sets of publications.
4. Lack of versatility: ensemble methods are effective in ML, but are not directly adapted to work with human expert assessments containing semantically rich categories.

Conclusion. Therefore, in modern literature, there is no universal approach that would allow:

- integrate multi-level expert classifications;
- take into account the relative weight of classes;
- maintain the interdisciplinarity of publications.

This confirms the relevance of developing a new method for coordinating expert assessments, which will be presented in this paper.

3. Methodology

The following procedure for processing and agreeing expert assessments in a generalized formulation is proposed. Suppose two experts have classified Y publications. Let us assume that the cardinality of the set of classes proposed to the expert is X, and the experts must assign each publication to no more than Xm ordered classes, with Xm<X. Then, two ratings of one publication are two vectors of classes defined by experts:

$$A =< a_1, a_2, ..., a_{X_m} >, B =< b_1, b_2, ..., b_{X_m} > \tag{1}$$

In each vector of evaluation of the publication belonging to a certain class, the positions of the vectors can contain elements from a set of predefined list of classes or an empty value ($\emptyset$). We believe that the assignment of the main class (vector positions – a_1, b_1) is mandatory, the following vector positions are filled only in the case when experts consider the publication to correspond to several classes. In this context $\emptyset$ means the absence of an expert assessment in the relevant position, not a separate class. In formulating a method for reconciling the results of expert classification of publications, the first step is to determine the distance metric between the two estimates.

When constructing a distance metric, we follow the axiomatic definitions of the metric:

- the distance between any two elements cannot be negative;
- the distance is zero only when the elements are the same;
- the distance from the first vector to the second is equal to the distance from the second vector to the first;
- triangle inequality - the straight line distance between two vectors is no greater than the circumnavigation through the third vector.

As an axiom, we also assume that the distance is equal to one if the vectors are completely different and do not intersect. In this case, in the case of complete disagreement or lack of intersection of the classification results from different experts, it will be possible to achieve at least some common understanding only through repeated expert evaluation.

As a metric of the distance D(A, B) between two expert estimates, we will take the minimum total cost of editing the vectors in such a way as to achieve their complete coincidence, similar to the Levenshtein distance. When editing, it is allowed to apply a predefined set of operations:

- $Replace_i$ – replace the class at position i in the vector;
- $Rotate_{i,j}$ – permutation of two classes in the score vector;
- Add_i – add a new class to the score vector, which can only be used in the case of an empty value ($\emptyset$).

We also set the following constraints on the cost of editing operations for arbitrary X_m (Table 1).

Table 1
Limit on the cost of editing operations at random X_m

Name of the operation	Limitation
Replace	$\forall i, j; i > j \Rightarrow Replace_i < Replace_j$ $\sum_{i=1}^{X_m} Replace_i = 1$
Rotate	$\forall i, j, k; k > i > j \Rightarrow Rotate_{j,k} \leq Rotate_{j,i} + Rotate_{i,k}$ $\forall i, j, k; k > i > j \Rightarrow Rotate_{i,k} < Rotate_{j,k}$ $\sum_{i=1}^{X_m-1} \sum_{j=i+1}^{X_m} Rotate_{i,j} = \frac{1}{2}$
Add	$Add_i = Replace_i$

Since the components of the publication classification vector are initially ordered according to their importance, the distance calculation algorithm can be characterized by the recurrent relation:

$$D_i = min\Big(D_{i-1} + Replace_i, D_{i-1} + Rotate_{i,k} | k > i, D_{i-1} + Add_i\Big) \tag{2}$$

where state Di is the minimum cost of editing the vector up to and including the i-th position, and $D_0 = 0$. If vectors A and B contain different numbers of empty values (∅), then the Add operation must be applied to the vector that has more empty values. Formulation of the problem of finding a compromise between mismatched expert estimates: for known values of two vectors A and B, it is necessary to find a vector S* such that the maximum distance from it to both input vectors is minimal:

$$S^*(A, B) = argmin\Big(max\Big(D(S, A), D(S, B)\Big)\Big) \tag{3}$$

This idea of finding the mean can also be applied if the evaluation is carried out by N experts: an example of the

$$S^*(A_1, A_2, ..., A_N) = argmin\Big(max_{i=1}^{N}\Big(D(S, A_i)\Big)\Big) \tag{4}$$

The proposed solution has drawbacks, it only partially automates the search for the average value and may require further expert coordination, since there remains a multivariate range of possible candidates for representing classes in the average vector. As an example of the need for further expert coordination, consider the following situation which has repeatedly arisen when processing expert assessments on the subject of the article.

1. Each publication was allowed to be assigned to no more than three classes. The first main class is mandatory.
2. The costs of editing operations are defined as follows:
    - $Replace_1 = 0.6, Replace_2 = 0.25, Replace_3 = 0.15;$
    - $Rotate_{1,2} = 0.2, Rotate_{1,3} = 0.25, Rotate_{2,3} = 0.05;$
    - $Add_2 = 0.25, Add_3 = 0.15.$

    Since the Add weights correspond to the Replace weights, a sensitivity analysis (SA) was conducted for Replace and Rotate using an "One at a Time" (OAT) perturbation, with the other weight vectors fixed as above for expert estimates (5). The datasets for the study were generated with a step size of 0.05 for each weight, and all the data sets generated satisfy the constraints given in Table 1. Based on the results of the SA, the defined cost indicators were selected from a small subset of the gold standard and can be considered generalizable to different expert groups.
3. The two peer reviews of the publication are as follows:
    $A = \langle Class_1, Class_2, Class_3\rangle, B = \langle Class_2, Class_1, Class_4\rangle.$

In this case, the distance between the two estimates is the sum of the values of the $Rotate_{1,2}$ and $Replace_3$ operations, which is 0.35. There are two alternative options for the mean vector S*, the distance from which to the input vectors will not exceed 0.2: $S_1^* = \langle Class_1, Class_2, Class_4\rangle$, $S_2^* = \langle Class_2, Class_1, Class_3\rangle$. If it is desirable to avoid a second examination and both experts have the same level of confidence in their competence, then the best solution is to choose the average vector randomly. For a more detailed analysis of this limitation, let us define the probability distribution function of the event that, as a result of independent classification of publications, experts selected q common classes. That is:

- There is an alphabet A, which includes X unique characters, namely a list of classes to which it is proposed to attribute each publication;
- expert evaluations are filled independently, without repetitions of symbols (classes) with a limited length $X_m < X$. These are two subsets where $S_1,\ S_2 \subset A, |S_1| = x_1, |S_2| = x_2, x_1, x_2 \in \{1,\ 2, \ldots, X_m\}$;
- The order of the characters does not matter (the probability of intersection depends only on the set of characters);
- all subsets of fillings are equally likely;
- the lengths of the filled lines are also equally likely with the restriction that the length of the line is randomly chosen from the range $1 \ldots X_m$.

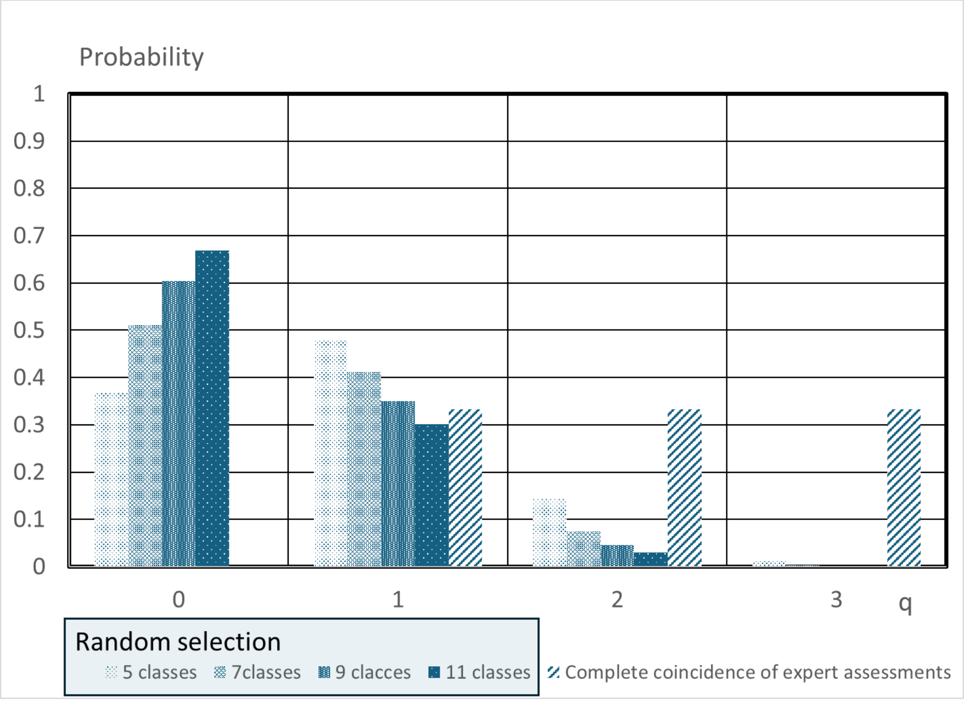

Figure 1: Comparison of the probabilities of random evaluation and expert evaluation with completely coincident evaluations with different numbers of classes proposed for classification ($X_m = 3$).

Then, with fixed $x_1, x_2 \in \{1,\ 2, \ldots, X_m\}$, the conditional probability of occurrence of q intersections is:

$$P\left(q, x_1, x_2, X\right) = \frac{\binom{x_1}{q}\binom{X-x_1}{x_2-q}}{\binom{x_2}{q}} \tag{5}$$

where $\binom{x}{y} = \frac{x!}{y!(x-y)!}$ is the number of combinations of selecting y elements from a set containing x elements.

After averaging over all $x_1, x_2 \in \{1, 2, \ldots, X_m\}$, taking into account that x_1, x_2 are random and uniformly distributed, we obtain a lower bound for the quantitative assessment of the coincidence of expert assessments. In essence, this is the probability of a random variable that describes the number of common elements (intersections) between two random subsets of the alphabet X, with random lengths uniformly distributed from one to X_m. In other words, this is the probability of a situation where, instead of assessing, experts indicated random classes:

$$\underline{P}\left(q, X_m, X\right) = \frac{1}{X_m^2} \sum_{x_1=1}^{X_m} \sum_{x_2=1}^{X_m} \frac{\binom{x_1}{q}\binom{X-x_1}{x_2-q}}{\binom{x_2}{q}}. \tag{6}$$

If the expert estimates completely coincide, the probability will be as follows:

$$\overline{P}\left(X_m\right) = \frac{1}{X_m}. \tag{7}$$

Figure 1 shows a comparison of the probabilities of randomly assigned scores for different numbers of classes proposed for selection – X, a fixed value $X_m = 3$, and fully agreed, congruent expert assessments. Formulas (6, 7) make it possible to calculate the mathematical expectation of the number of intersections in expert estimates (Table 2):

$$\underline{Q}\left(X_m, X\right) = \sum_{q=1}^{X_m} \underline{P}\left(q, X_m, X\right) \cdot q, \tag{8}$$

$$\overline{Q}\left(X_m\right) = \sum_{q=1}^{X_m} \overline{P}\left(q\right) \cdot q = \frac{X_m + 1}{2}. \tag{9}$$

The normalized metric of expert assessment quality $Q(X_m, X)$ is constructed based on the assumption that the minimum value $\underline{Q} = 0$ corresponds to random assignment of classes, the maximum value $\overline{Q} = 1$ can be achieved only with complete coincidence of expert assessments. A negative value of Q is possible if experts, for some reason, intentionally indicated incorrect answers.

Then the quality of the assessment through the metric of coincidence of expert assessments is determined as follows:

$$Q(X_m, X) = \frac{Q_f\left(X_m, X\right) - \underline{Q}\left(X_m, X\right)}{\overline{Q}\left(X_m\right) - \underline{Q}\left(X_m, X\right)}, \tag{10}$$

where $Q_{av}(X_m, X)$ – the calculated average value of the number of intersections in expert assessments of the membership of Y publications to X classes, with the restriction that each publication can be assigned to no more than X_m classes.

$$Q_{av}\left(X_m, X\right) = \frac{\sum_{y=1}^{Y} t_y}{Y}, \tag{11}$$

where t_y is the number of common classes in two expert assessments of one publication.

Table 2
Calculation results $\underline{Q}\left(X_m, X\right), \overline{Q}\left(X_m, X\right)$, at different values X_m, X

		X_m		
	X	1	2	3
$\underline{Q}(X_m, X)$	5	0.20	0.45	0.80
	7	0.14	0.32	0.57
	9	0.11	0.25	0.44
	11	0.09	0.20	0.36
$\overline{Q}(X_m)$		1.00	1.50	2.00

Thus, firstly, the lower and upper limits for assessing the consistency of expert evaluation results are determined, and, as a result, a normalized metric of the quality of expert evaluation is proposed. Secondly, a method is proposed to coordinate expert assessments and find a compromise.

4. Experimental Results

4.1. Data Set

To test the proposed method, a sample of 203 scientific publications devoted to the architecture of onboard computers (OBC) for CubeSat nanosatellites was formed [2]. The data sources were the electronic libraries IEEE Xplore and Scopus in the period 2015–2024. The PRISMA [9] methodology was used to select articles, which included:

- primary search by keywords (“CubeSat”, “Onboard Computer”, “Fault Tolerance”, “Software Architecture”, etc.);
- removing duplicates and irrelevant publications;
- verification of compliance with inclusion criteria.

The result was a representative sample of works covering both hardware and software aspects of building a CubeSat OBC.

4.2. Experts

Two experts were involved in the classification process:

- **Expert 1** — a specialist in the industrial IT industry, specializing in building embedded real-time systems.
- **Expert 2** — a representative of the university environment who researches on-board systems architectures in scientific projects.

Each expert classified all 203 publications according to a defined system of categories. For each article, experts were required to indicate a main class and up to two additional classes.

4.3. Classification System

Six main classes have been identified, reflecting key areas of CubeSat OBC research, and they are listed below.

1. Hardware Architectures.
2. Software Systems.
3. Fault Tolerance & Reliability.
4. Power & Resource Optimization.
5. Communications and interfaces (Communication & I/O).
6. Intelligent functions (AI & Autonomy).

Each publication could be assigned to several classes simultaneously, reflecting its interdisciplinary nature.

4.4. Comparable Methods

Three approaches were compared to assess effectiveness:

1. Majority voting — choosing the class that is most frequently found among expert assessments.
2. Cohen's Kappa — assessment of the consistency of two experts without building an integrated classification.
3. Proposed method — distance metric + search for a compromise solution.

4.5. Results

- According to the majority voting method, 57 publications (28%) turned out to be conflicting, where experts chose different main classes;
- According to Cohen's Kappa coefficient, the level of agreement was 0.62 (average agreement);
- The proposed method allowed for the agreement of most cases, leaving only 18 publications (9%) that required re-examination due to too high a distance between the assessments.

In addition, the method was particularly useful in cases of interdisciplinary articles. For example, if one paper described both hardware solutions and fault tolerance algorithms, traditional methods "forced" the selection of only one class. Instead, our method allowed us to maintain a multi-level evaluation structure (main + additional classes), which provided a more objective reflection of the content.

5. Discussion

The results obtained show that the proposed method of harmonizing expert classifications of publications has a number of advantages compared to traditional approaches:

- reduces the number of conflicts between experts by more than three times (from 57 to 18 publications);
- provides a better interpretation of results in the case of interdisciplinarity;
- allows you to quantify the distance between assessments and use this metric as a criterion for re-examination.

Thus, the method has proven its effectiveness as a tool for harmonizing expert classifications, combining formalism with practical applicability.

5.1. Advantages of the Method

1. Preservation of multi-level classifications. Unlike most existing approaches, the method does not reduce the assessment to just one "correct" class, but allows for a hierarchical structure (main class + additional ones). This is especially important in interdisciplinary research, where one article can simultaneously belong to several directions.
2. Quantifying differences. Using the distance metric allows us to calculate the "degree of divergence" between expert assessments. This allows us to more objectively determine when the difference is insignificant (for example, a permutation of classes) and when it is critical (replacement with a completely different class).
3. Flexibility in customization. The cost of Replace, Rotate, and Add operations can be tailored to specific tasks, giving more or less weight to individual aspects of the classification. This makes the method versatile for different subject areas.
4. Reducing the number of conflict cases. In the experiment, the number of publications requiring re-examination decreased from 57 (28%) to 18 (9%). This indicates an increase in the efficiency of the systematic review process.

5.2. Limitations

Despite the obvious advantages, the method also has certain limitations:

- dependence on the choice of weighting factors. If the values of operations (Replace, Rotate, and Add) are set incorrectly, the result may be skewed. Therefore, it is necessary to calibrate the parameters for each subject area;
- a small number of experts in the study. The presented experiment involved only two experts. For greater reliability, tests with a wider group of evaluators are needed;
- the need for manual intervention in complex cases. If the distance between the ratings exceeds a certain threshold, the article still needs to be re-examined. It is currently impossible to completely eliminate the human factor.

5.3. Potential Areas of Application

The proposed method can be used not only for classifying publications, but also in a wider range of tasks:

- bibliometrics and scientometrics: construction of systematic reviews, formation of taxonomies of scientific areas, analysis of interdisciplinary research;
- peer review and expert evaluation: coordination of decisions of multiple reviewers when evaluating articles or grant applications;
- medical diagnostic systems: combining the conclusions of several expert doctors to form a more reliable diagnosis;
- decision support systems: use in project management, multi-criteria evaluations, or collective strategy selection;
- artificial intelligence and machine learning: adaptation of the method as a tool for building ensembles in classification problems, where various algorithms act as "experts".

5.4. Generalization

Thus, the proposed method can be considered as a hybrid approach that combines the ideas of classical consistency metrics (Kappa, Kendall’s W) with the concepts of ensemble learning. Its main advantage lies in the ability to work with ordered sets of classes and in preserving the multilayer structure of estimates, which is especially relevant for the analysis of complex interdisciplinary data.

6. Conclusions

In this paper, a method for harmonizing the results of expert classification of scientific publications was proposed and tested, which combines approaches from bibliometrics and artificial intelligence. Unlike traditional methods, such as majority voting or consistency coefficients (Cohen’s Kappa, Kendall’s W), the developed approach takes into account the multi-level structure of classifications (main and additional classes) and allows for quantitative assessment of discrepancies between expert assessments.

The proposed method is based on the introduction of a distance metric between expert assessments, which takes into account the operations of replacement (Replace), rotation (Rotate), and addition (Add). This makes it possible to form a compromise solution that minimizes the distance to the assessments of all experts, as well as to identify publications that require re-examination. The method was tested on a corpus of 203 publications on the architecture of CubeSat onboard computers. Two experts with different professional experiences participated in the study. The results showed that the number of conflict cases decreased more than threefold (from 57 to 18 publications), and the level of classification stability increased significantly. The main conclusions of the study can be formulated as follows.

1. The proposed method provides a more objective and interpretable agreement of expert assessments.
2. It allows you to maintain the interdisciplinary nature of classifications without reducing publications to a single class.
3. The use of a quantitative distance metric opens up opportunities for automating the matching process.

Further research directions:

- expanding the method to work with a larger number of experts;
- integration of weighting factors depending on the expert’s qualifications or industry specifics;
- application in the field of peer review, medical expert systems, and other domains where the integration of the opinions of several specialists is important;
- development of machine learning-based tools for automated reconciliation support.

Thus, the developed approach contributes to the development of the methodology for analyzing scientific publications and can become an effective tool for improving the quality of systematic reviews in various fields of knowledge.

7. Authors’ Contribution

Problem formulation – I. Turkin; review and analysis of information sources – A. Chukhray; development of a method for processing and coordinating expert assessments – O. Liubimov, L. Volobuieva; conducting research, evaluation, and visualization of results – I. Turkin, A. Chukhray.

Acknowledgments

The study was funded by the National Research Foundation of Ukraine in the framework of the research project NRFU.2023.04/0143 - "Experimental development and validation of the on-board computer of a dual-purpose unmanned aerial vehicle". The authors express their gratitude to the graduate student of the Department of Software Engineering of the National Aerospace University, Valkovy VS, for his work as an expert in the classification of scientific publications.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] J. Cohen, A coefficient of agreement for nominal scales, Educational and Psychological Measurement 20 (1960) 37–46.

[2] I. Turkin, O. Liubimov, L. Volobuieva, V. Valkovyi, Hardware and software of cubesat nanosatellites' on-board computers: a systematized literature review, Aerospace Technic and Technology 0 (2025) 116–137. doi:https://doi.org/10.32620/aktt.2025.3.11.

[3] M. G. Kendall, Rank Correlation Methods, Griffin, 1970.

[4] T. G. Dietterich, Ensemble methods in machine learning, in: Multiple Classifier Systems, Springer, Berlin, Heidelberg, 2000, pp. 1–15. doi:10.1007/3-540-45014-9_1.

[5] Z.-H. Zhou, Ensemble Methods: Foundations and Algorithms, CRC Press, 2012.

[6] A. Strehl, J. Ghosh, Cluster ensembles—a knowledge reuse framework for combining multiple partitions, Journal of Machine Learning Research 3 (2002) 583–617.

[7] H. Linstone, M. Turoff, The Delphi Method, Addison-Wesley, 2002.

[8] T. L. Saaty, Decision making with the analytic hierarchy process, Int. J. Services Sciences 1 (2008) 83–98. doi:10.1504/IJSSCI.2008.017590.

[9] PRISMA, PRISMA statement, https://www.prisma-statement.org/, 2020. Accessed on 2025-09-03.