ArchiBERTo: a Hierarchization Quality Objectives NLP Tool in the Italian Architecture, Engineering and Construction Sector

Mirko Locatelli [1], Lavinia C. Tagliabue [2] and Giuseppe M. Di Giuda [3]

[1] Politecnico di Milano, Department of Architecture, Built Environment and Construction Engineering, 20133 Milan, Italy
[2] Università degli Studi di Torino, Department of Computer Science, 10149 Turin, Italy
[3] Università degli Studi di Torino Department of Management, 10134 Turin, Italy

Abstract
Natural language is the main source of communication during pre-design phase, effective communication among the actors must be guaranteed for the design project success during this crucial phase. In the proposed study, textual data is processed via an NLP tool (ArchiBERTo) specifically developed for the elaboration of Design Guidance Documents (DIP), pivotal documents in the pre-design stages of the design and construction procurement process in Italy. DIP defines demands and objectives of the public appointing party. The tool is used to process and translate the DIP quality objectives related sentences into a list of hierarchized objectives and criteria. To evaluate ArchiBERTo performances, the outputs generated by the tool and the objectives rankings provided by a group of architecture and construction experts are compared. The results show a good capability of the tool to mirror the collective capability and sensitivity of the group of experts in the design and construction domain.

Keywords
Collective intelligence, BERT, School buildings

1. Introduction

1.1. Natural language and pre-design phase in Architecture, Engineering and Construction sector

Pre-design is the initial and a crucial phase of the architectural design and construction process having a significant impact on the project's value [1]. During the pre-design phase, project goals and objectives are defined and conveyed to the designers in order to reach a consensus between the stakeholders' needs and demands and the designers' proposals. Effective communication and the consequent proper understanding of requests and requirements by all the involved parties is the main goal of the pre-design phase [2], being the objective definition, communication, and understanding a critical factor for the success of the design and construction projects [3]. In the pre-design phase, communication mainly takes place using verbal expressions collected and shared through multiple text documents [4], and natural language turns out to be the main source of information at this stage of the design and construction process. However, natural language can lead to misinterpretation, or at least different interpretations and complexities [5], primarily in the definition of the relative importance of the quality needs and objectives to be pursued in the project. In fact, the hierarchy assigned to demands, especially qualitative demands, varies greatly from subject to subject being a personal and individual judgment influenced by countless factors and biases. In addition, considering the prohibition of direct communication for the actors involved in a public call for tenders, the forementioned obstacles inherent in the use of natural language turn out to be exacerbated [6].

1.2. Italian public design call for tenders: actors, documents, and procedure

An overview of the design call for tenders procedure and of the actors' role involved within the Italian design and construction sector is presented to better identify information flow criticalities linked with the mandatory steps of the tender procedure. In the following sections the research context and objectives are also introduced. As stated, the procedure of a public design call for tender, in the Italian context, involves mandatory steps, documents and the participation of three main actors:

- Appointing public party: it identifies needs, objectives, and requirements to be pursued by the design project. Quality and quantitative objectives are defined and shared via a Design Guidance Document called Documento di Indirizzo alla Progettazione (DIP).
- Design teams: teams of designers competing in the call for tenders. To win the tender they aim to submit a design proposal that meets the public actor's requirements and needs as defined in the DIP.
- External committee: appointed by the appointing public party, the committee evaluates the design teams bids to identify the best design project, i.e., the one that comprehensively complies the systemic demands and requirements declared as priorities, via natural language expressions, in the DIP document.

From this point of view, the DIP is the instrument for the public party to communicate quality demands and expectations regarding the design and construction of the building and, at the same time, the benchmark adopted to evaluate the design proposals submitted by the competing design teams [6]. Consequently, the DIP aims to ensure that the interventions will meet the administration needs and objectives which must be clearly identified and stated in the document. In order to reach this goal, the DIP should allow the designers to have a deep understanding about the needs and objectives which ought to be properly communicated and shared to lead the design proposals towards the achievement of the correct goals, and at the same time, support the external committee in the evaluation of the bids, defining the quality objectives relative priority and hierarchy. Moreover, the DIP contents are regulated in the Italian legislation by the D.P.R. 207/2010, the document is divided into two main sections. A quantitative section about the state of the premises, technical requirements, and regulations. This section can be defined and supported identifying alpha-numerical parameters; a qualitative section that describes quality objectives and expectations (e.g., sociocultural value of the project, architectural and landscape quality of the intervention, flexibility of spaces, perceptual comfort, etc.).

Regarding the quantitative section architects and engineers are already prepared and equipped with specific digital tools and calculation methods to deal with alpha-numerical information mainly using the Building Information Modelling (BIM) approach. BIM is a methodology to digitally manage the design and construction process allowing to model and represent a physical asset, like a building or building components, in a virtual environment [7]; BIM methodology and related enabling tools have been used as a design management approach by the design and construction industry in order to improve the collaboration and communication among the construction players as well as the management of documentation in the construction projects, helping to accomplish efficiency and effectiveness [8]. Consequently, the evaluation of the design proposals by the external committee on the standards and technical requirements aspects is currently feasible applying and calling for the application of BIM methodology, since it is based on processing and comparing procedures implying parameters and numerical values. From this point of view, the evaluation of technical needs can be supported by requiring the designers to deliver specific Building Information Models. On the contrary, being the DIP qualitative section expressed and shared via natural language expressions, the processing and digital management of quality characteristics cannot be handled via traditional BIM methods and digital tools.

1.3. Natural Language Processing in Architecture, Engineering and Construction sector

As explained, a major part of the quality-related information is expressed relying on natural sentences and exchanged via written text documents. Unstructured data and information, such as written natural language, can be processed and digitally managed relying on Natural Language Processing

(NLP) systems and techniques. NLP aims to allow computers to process human natural language and knowledge [9–13]. NLP systems have been already applied and assessed in several AECO sector fields [14]. A brief overview of the existing NLP applications in AECO is provided to frame existing studies, detect possible deficiencies, and research gaps.

Articles about NLP applications in AECO are listed according to major application fields (i.e., Procurement management, Safety management, and Project and construction risk management) and scope/task:

- Procurement management: Legal clauses classification [15,16]; Contract document risk detection [17–19]; Automated detection of contract changes [20]; Disputes resolution facilitation [21].
- Safety management: Safety risks prediction [22,23]; Construction site accidents analysis [24]; Safety incompatibilities prediction [25]; Accidents and injuries prediction [26–28].
- Project and construction risk management (different from safety and legal risks): Requirements defects detection [29]; Estimation of non-compliance [30–33]; Support project and construction risk management [34]; Support or automate compliance checking [35,36].

The analysis highlights the lack of applications involving documents belonging to the pre-design or preliminary stages of the design and construction process. For a detailed and in-depth scientometric analysis of the use of NLP and BIM in AECO sector the authors suggest the consultation of Locatelli et al. [37].

1.4. NLP tool for automatic classification of qualitative objectives

The study aims, by the use of an ad-hoc developed NLP tool (ArchiBERTo) [6,38], to automatically process and translate the quality objectives expressed in DIPs into a list of hierarchized objectives to support the evaluation process of the design proposals of school buildings. The use of ArchiBERTo aims to establish a consensus between the actors regarding the relative hierarchy of quality needs and objectives, as shown in Figure 1. The manuscript explains the development and assessment steps adopted and the evaluation of the tool measuring the subjectivity degree and customization capability, fundamental features to properly translate the quality objectives and needs related sentences into a list of hierarchized objectives and criteria. As an ultimate goal the use of ArchiBERTo aims to minimize the possible different interpretations of the hierarchy of the appointing party objectives by the design teams and by the external committee enhancing the effective communication during the pre-design phase.
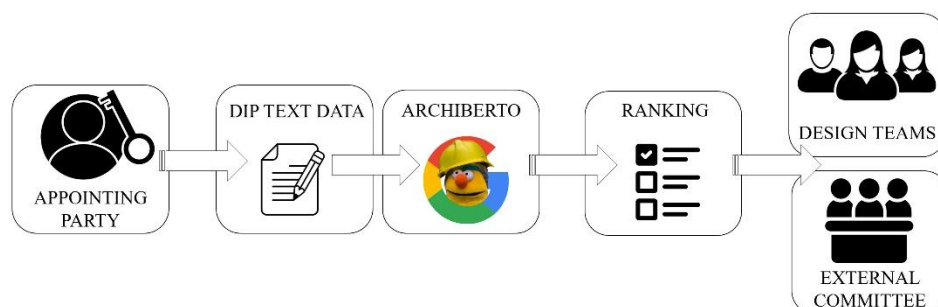

Figure 1: Schema of ArchiBERTo DIP processing.

2. Methodology

For reasons of clarity the methodology section is divided into two main parts. The first one explains the development and assessment steps adopted to train and develop the model, also providing the metrics used to measure the ArchiBERTo performances. The second focuses on evaluating the NLP

tool by calculating the degree of subjectivity in objectives and criteria ranking generation and assessing the capability to produce goal hierarchies customized on the content of the processed documents.

2.1. ArchiBERTo development and assessment steps

The NLP-based tool (ArchiBERTo) is developed as a multi label classifier based on the BERT (Bidirectional Encoder Representation from Transformers) language representation model that provides contextualized embedding [39]. A multi-label classifier has been chosen because capable to automatically apply more than one classification label to a single text or sentence allowing the prediction of multiple mutually non-exclusive classes [40], classes which coincide with a defined list of labels. The capability to automatically label and according to the labels assign weights is the basis to generate the priority ranking of the quality objectives of the DIPs used as case study. The NLP tool is trained (using the data from already validated DIPs) to classify sentences according to the set of predefined labels which represent the appointing party and end-users' demands and quality requests. The main activities to develop and assess the tool are listed below and explained in the following paragraphs:
- labels definition
- training and validation dataset production
- model fine-tuning
- performance evaluation

2.1.1. Labels definition

As stated, the NLP tool is trained to classify sentences according to a set of predefined labels. In this phase is fundamental to create a consensus about the labels number and definition, representing the appointing party and end-user interests and quality objectives. Consequently, to reach a consensus among the stakeholders involved the set of labels must be defined in conjunction by the appointing party, the end-users, when possible, and the domain experts; in this specific context architects, building engineers, and designers. For the selected case study (Progetto Iscol@), a list of predefined labels, defined by the appointing party, is already available. The labels are the result of a cooperation among different experts (i.e., architects, designers, pedagogues, and agronomists) and end-users (i.e., primary and secondary school teachers, principals, and school building janitors) representing the result of a collective effort of the different stakeholders.

2.1.2. Training and validation datasets definition

Being the proposed NLP tool based on the BERT language model it is necessary to fine-tune the model to solve the multi-label classification problem in the architecture and design knowledge domain. Consequently, a certain amount of training and validation data is needed. A general dataset is defined and then is randomly split into a training and validation dataset at a 0.8:0.2 ratio. The general dataset is defined by selecting DIP sentences and manually assigning labels.

The manual labelling is a critical task influencing both accuracy and capability of the NLP tool to automatically process and properly label needs and quality objective sentences. Being the dataset the knowledge source of the tool the general dataset is produced via a collaboration between experts with knowledge in the architectural, design, and construction fields. In addition, since the proposed NLP system is applied to a specific case study (Progetto Iscol@) a deep knowledge of the strategic objectives of Progetto Iscol@ are requested to the experts to correctly label the training sentences. Since Iscol@ members could not be directly involved in the project, a preliminary study of the overall goals, guidelines, and context of Iscol@ is conducted by the selected experts before the labelling of the training sentences. Moreover, to avoid biases in the production of the dataset, each expert is asked to independently propose a hypothesis for the labelling of each sentence. Then the experts share their hypothesis and in case of disagreement on some labels, they are asked to share the motivation of their

label choices and converge on a single common proposal. The construction of the dataset by different experts aims to allow the model to represent and use their collective knowledge in the labelling activity.

The NLP tool aims to avoid subjectivity in the interpretation of textual information by representing the collective intelligence of a group of experts, rather than that of a single expert. Furthermore, ArchiBERTo aims to outperform the capability of a single expert to manage the complexity of analyzing several sentences being the representation of a group of experts' knowledge.

2.1.3. Model fine-tuning parameters

Once defined the dataset, to properly train the BERT model, a set of hyperparameters must be defined. A hyperparameter is a variable configuration that is external to the model and whose value is not estimated from the data but estimated via a trial and errors cycle. The list and the description of the hyperparameters used for the NLP tool training is provided in Table 1.

Table 1
Hyperparameters description and values setting.

Hyperparameter	Description
MaximumLength	Maximum number of words simultaneously processed during the training
TrainingBatchSize	Number of training examples used in one iteration
ValidationBatchSize	Number of examples used for the validation in one iteration
EpochsNumber	An epoch is an entire transit of the training data through the algorithm
LearningRate	It defines the adjustment in the weights of the neural network with respect to the loss gradient descent

2.1.4. Performance assessment metrics and learning curves

In order to measure the NLP tool accuracy, the model predictions are compared with the human annotation of the validation dataset. Precision (P), Recall (R) and F1-score (F1) metrics are selected to measure the model performances [41].

Learning curves are also plotted, in fact learning curves display the training error as a function of the number of iterations in the optimization process allowing the monitoring of the optimality of a model allowing to diagnose problems and optimize predictions [42]. Specifically, training and validation loss curves are plotted in order to detect the properly training of the model identifying possible underfitting or overfitting behaviors.

An underfitted model shows a loss value function for training and validation curves not decreasing with the number of iterations or epochs. An underfitted model is highly biased, and it does not consider the data and relevant information. On the other hand, an overfitted model shows a decreasing training loss curve, achieving low error values per iterations or epochs. However, in an overfitted model the validation loss decreases until a minimum turning point and then starts to increase. The minimum point represents the beginning of the overfitting behavior of the model. If overfitted the model can capture and learn from training data, but it performs poorly on new and unseen data, showing poor model generalization performances. Consequently, to proper train a model is necessary to stop the training process at the global minimum point, i.e., where the validation error trend changes from descending to ascending. Summarizing, if the training process is stopped before the global minimum point the model is underfitted, if it is stopped after the global minimum point the model is overfitted, as shown in Figure 2.
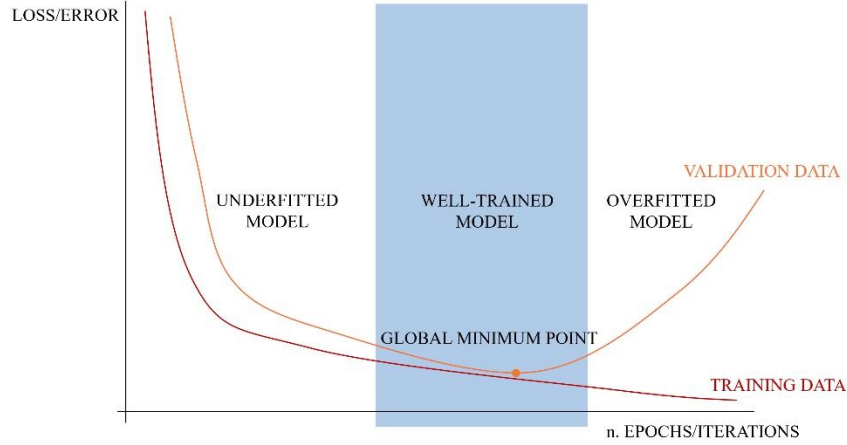
Figure 2: Underfitting, overfitting and optimal training zone.

2.1.5. ArchiBERTo outputs and ranking calculation

Once ArchiBERTo is fine-tuned and the performances are assessed, the NLP tool can be tested processing new sentences assigning the labels and the accuracy degree with which the labels are associated to the new sentences. The accuracy degree values of each processed sentence represent the weights of the labels, and thus the relative priority of the labels/quality objectives for the single sentence. The accuracy values/weights of the labels obtained by processing all the sentences of a document are summed and normalized to define the total weight of each label for the entire DIP (Formula (1)).

$$Label\ Weight_i\ = \left. \Sigma\ (L_i) \middle/ \Sigma\ (L_1) + \Sigma\ (L_2) + \Sigma\ (L_3) + \cdots + \Sigma\ (L_n) \right. \tag{1}$$

where *Li* denotes the accuracy value of the *i-th* label, *i = (A.1, B.1, C.1…, P.1)*

The total weights of each label represent the relative importance of the quality objectives to be pursued by the design teams in the definition of the design proposals, and at the same time, the evaluation criteria to be used by the external committee to evaluate the design proposals.

2.2. ArchiBERTo evaluation
2.2.1. ArchiBERTo subjectivity degree

As stated, after the completion of the NLP tool development and assessment the evaluation of the subjectivity degree of the tool is conducted. In order to evaluate ArchiBERTo subjectivity the contents of different DIPs related to the quality objectives are processed by ArchiBERTo and the corresponding ranking of objectives/criteria is provided.

The same DIPs are also analyzed by three experts individually. Each of them individually hierarchize the objectives providing a ranking, then the same DIPs are read and analyzed by the three experts collectively and a ranking is provided by the group.

To measure the subjectivity degree of the tool, the rankings generated via the NLP model and the rankings provided by the single experts are compared with the rankings provided collectively by the group of three experts considered as the benchmark. To calculate the discrepancy between the evaluations of the individual opinions of each expert and the NLP tool with the collective evaluation, a score is assigned from 1 (last goal/label in the ranking) to 21 (first and most important objective in the ranking). The discrepancy coincides with the difference between the score/position of each objective generated by each single expert and by the NLP tool, both compared with the ranking provided collectively by the group of experts (Figure 3).

Figure 3: Single experts and ArchiBERTo ranking subjectivity measurement.

Consequently, two conditions can happen:

- ArchiBERTo ranking discrepancy for each label > individual expert ranking discrepancy for each label: the tool has a higher degree of subjectivity than the single experts and for that reason does not adequately represent the collective capability to translate quality objectives expressed in natural language into the corresponding ranking.
- ArchiBERTo ranking discrepancy for each label < individual expert ranking discrepancy for each label: the tool is less affected by subjective biases representing the collective capability of the group of experts to translate the natural language expressions into a ranking of objectives.

2.2.2. ArchiBERTo customization degree

Three different DIPs are processed, and three rankings are produced using the NLP system. A unique ranking is produced by a group of three experts from the analysis of a single DIP. The aim is to measure the capability of the NLP system to generate a ranking customized on the DIPs content, measuring the variation of the rankings of different DIPs compared with the fixed ranking produced by the group of experts collectively (Figure 4). The described experiment aims to measure the flexibility of the proposed system, which must not flatten the rankings of objectives proving the capability of ArchiBERTo to provide a customized prioritization of objectives for different DIPs, mirroring the semantic content of each document.
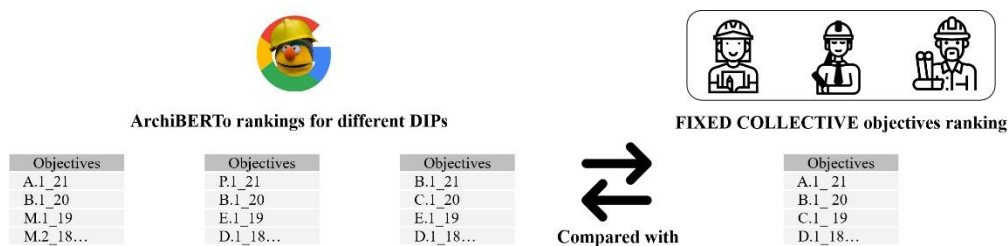

Figure 4: NLP tool ranking customization capability measurement.

3. Case study
3.1. Case study: Progetto Iscol@

The case study concerns the processing of DIPs regarding the design and construction of school buildings. In particular, Progetto Iscol@ introduced by the Regional Council of Sardinia to address the backwardness of the regional education system, aiming at modernizing and expanding the regional school building stock, is chosen as a suitable case study. The public investment aims at achieving high standards of architectural quality and social and environmental sustainability of the educational facilities. General directions and regional guidelines are shared in the early stages of Progetto Iscol@ to the various municipalities involved in the interventions. The use of guidelines ensures that all DIPs follow the regional directives, homogenizing the quality objectives of interventions on the island's school heritage through the sharing of a list and a ranking of quality objectives defined by the Iscol@

working group. The completion of the first round of call for tenders and pilot projects highlighted the impact of using a standardized list and ranking of objectives with a relative priority fixed for all the projects. On one hand, it demonstrated to be an effective approach for leading different projects toward criteria and objectives in line with Iscol@ strategic goals. On the other hand, it turned out to be an excessively rigid approach to properly support the designers and the external committee in the evaluation of school building projects that differ in geographical-environmental, socio-cultural context and consequently in quality demands. In fact, considering the internal specificity of design and construction projects, as they are closely related to and influenced by the context and by the different socio-economic and territorial needs, the use of a fixed hierarchy of objectives tends to flatten and eliminate the individual specificities and needs of the different projects. Consequently, the use of NLP can, in such a context, play a crucial role customizing the prioritization of objectives for each call, mirroring the semantic content of each DIP. Consequently, the use of the ad-hoc developed NLP tool ArchiBERTo, aims to reintroduce proper flexibility and compliance with the different project specific demands and requests.

4. Results and discussion

4.1. ArchiBERTo development and assessment steps

As stated, the NLP-based tool ArchiBERTo is developed as a multi label classifier based on the BERT (Bidirectional Encoder Representation from Transformers) language representation model. Results and details of each step of the development and assessment of the tool are provided in the following sections, followed by a section about the result of the tool evaluation (i.e., subjectivity degree and customization capability).

4.1.1. Dataset construction

As stated, a BERT language model is fine-tuned to solve the multi-label classification problem in the architecture and construction knowledge domain. Among the DIPs currently validated and published by the municipalities, twenty-one DIPs are used to develop and assess the NLP tool. From the DIPs collection a data set is produced and split (with a ratio of 0.8:0.2) into a training and validation dataset. Table 2 shows labels topic and the number of sentences manually labelled.

The dataset is available at the following link in .csv format: Github_ArchiBERTo_dataset.

Table 2
Training and validation dataset overview.

Label	Label/Objective topic	Sentences
A.1)	Capability of the school building to be used as a Civic Center	133
B.1)	Visibility and integration of sustainable design choices (educational medium) and integration of the intervention into nature and application of landscape enhancement strategies	45
C.1)	Possibility of personalization of spaces and equipment to prevent vandalism creating a feeling of belonging in users	25
D.1)	Spatial and volumetric integration of the intervention in the context and with existing buildings (shape, materials, colors, connections etc.) and proper mediation with the demand for visibility and architectural quality of the intervention as a building containing public functions	68
E.1)	Articulation of spaces and accesses with a focus on simple and clear identification of the various functions, including using colors and signages	61
E.2)	Presence of green spaces as an integral part of the design	34
F.1)	Perceptual quality (natural and artificial light) and psychophysical comfort (visual, thermo-hygrometric, acoustic etc.) to promote comfort and learning	144
F.2)	Indoor air quality and healthiness	29

G.1)	Cleanability, durability, maintainability, and replaceability of landscaping, materials, and greenery to reduce operating and maintenance costs	41
I.1)	Integration of the intervention with the road system and distinction between driveways, bicycle, and pedestrian paths; provision of areas and equipment to encourage slow and non-motorized mobility	36
I.2)	Ensuring accessibility and usability for people with disabilities	40
L.1)	Fostering interactions between students and teachers, group work and peer learning (collaborative learning and peer tutoring) by supporting innovative and inclusive teaching. Architecture should support the idea of space as a "third teacher"	197
L.2)	Visual and spatial continuity between outdoor (green and non-green) and indoor environments to encourage outdoor educational activities and enhance contact with the natural environment (outdoor space can be used as a second classroom). Connection between classroom and circulation spaces. The architecture should support the concept of openness of the traditional classroom and the concept of learning landscape	108
M.1)	Use of renewable, natural (non-harmful), local materials or materials with recycled content	46
M.2)	Minimization of the impact of the building on the surrounding environment (noise, light, water pollution, heat island effect, minimization of land consumption and use of soil defense strategies etc.)	89
M.3)	Integration between design and renewable energy production systems and exploitation/management of solar, light, and natural cooling and heating inputs	48
M.4)	Requests regarding energy standards and minimization of consumption (energy, water etc.) including using monitoring systems	96
N.1)	Ensuring safety during school activities and separation between activity conducted by people not belonging to the school staff, maintenance activities (spaces and paths). Adequate delimitation of the school perimeter, and need for control/supervision	26
O.1)	Spatial flexibility (furniture, facilities etc.)	198
O.2)	Temporal flexibility, possibility of use during curricular and extracurricular hours by citizens and long-term temporal flexibility, adaptability of spaces (readiness for change, adaptability)	103
P.1)	Usability of technological devices and integration with learning theories. Integration of space and technology; widespread presence of ICT technologies	103

4.1.2. Model fine tuning parameters

Description and values of hyperparameters used for the NLP fine tuning are provide in Table 3.

Table 3
Hyperparameters description and values setting.

Hyperparameter	Description	Value
MaximumLength	Maximum number of words simultaneously processed during the training	85
TrainingBatchSize	Number of training examples used in one iteration	2
ValidationBatchSize	Number of examples used for the validation in one iteration	32
EpochsNumber	An epoch is an entire transit of the training data through the algorithm	20

LearningRate	It defines the adjustment in the weights of the neural network with respect to the loss gradient descent	2 E-05

4.1.3. Fine-tuned model evaluation

As stated, to evaluate the NLP tool predictions on the validation dataset Precision (P), Recall (R) and F1-score (F1) metrics are calculated (Table 4) to measure the model performances.

Table 4
Model Precision, Recall, and F1-score values per label.

Label	Recall (R)	Precision (P)	F1-score (F1)
A.1)	0.92	0.71	0.80
B.1)	0.75	0.75	0.75
C.1)	1.00	0.75	0.86
D.1)	0.42	0.56	0.48
E.1)	0.64	0.88	0.74
E.2)	0.33	0.67	0.44
F.1)	0.77	0.68	0.72
F.2)	1.00	1.00	1.00
G.1)	0.90	1.00	0.95
I.1)	0.67	0.67	0.67
I.2)	0.86	0.55	0.67
L.1)	0.8	0.67	0.73
L.2)	0.39	0.85	0.54
M.1)	1.00	0.67	0.80
M.2)	0.90	0.82	0.86
M.3)	0.86	0.67	0.75
M.4)	0.60	0.64	0.62
N.1)	0.50	1.00	0.67
O.1)	0.71	0.88	0.78
O.2)	0.78	0.44	0.56
P.1)	1.00	0.65	0.79

The results show that the NLP model is properly fine-tuned showing only the labels D.1) and E.2) with a F1-score lower than 0.5.

Furthermore, training and validation loss learning curves are plotted to avoid overfitting or underfitting phenomena (Figure 5).
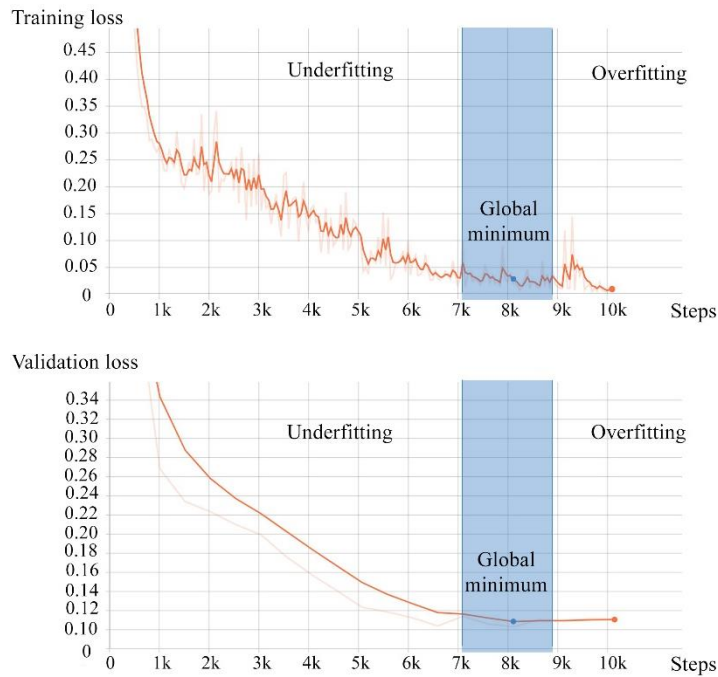
Figure 5: Training and validation loss charts.

The curves plot shows the training and validation loss curves gradually decrease and flatten moving close to each other, furthermore validation loss is slightly greater than the training loss at the global minimum point: Validation_loss = 0.1031, Training_loss = 0.08937. Consequently, the model can be considered fine-tuned. No overfitting or underfitting phenomena are present.

The code is available in .ipynb at the link: Github_ArchiBERTo_code.

4.1.4. ArchiBERTo output example

After the assessment and fine-tuning phase, a qualitative evaluation of the capability of ArchiBERTo in the labelling of new unknown sentences is conducted. An example of the processing and labelling output of two quality related sentences is provided in Figure 6.

Figure 6: Outputs of the NLP tool processing and labelling of DIP sentences.

4.2.   ArchiBERTo evaluation: subjectivity degree

In the the following sections the results of the subjectivity and customization degree evaluation of the tool are provided and discussed to demonstrate the capability of ArchiBERTo to mirror the collective intelligence and its capability to customize the objectives rankings according to the different DIPs content.

4.2.1. DIP number 1: Sassari primary school

ArchiBERTo is applied to process a DIP of a primary school located in the municipality of Sassari, Sardinia. The objectives ranking generated by the NLP tool gets for 12 out of 21 labels/objectives the lowest discrepancy. 4 labels show a discrepancy lower than the discrepancy of the evaluation of two expert, and other 4 labels, B.1), D.1), F.1) and O.2), show a discrepancy lower than a single expert and for only 1 objective, A.1), the NLP tool obtains the highest discrepancy compared with the benchmark. Moreover, considering the discrepancy (d) of the ranking provided by the NLP tool:

- 15 labels show $d < 10\%$
- 4 labels show $10\% < d < 20\%$
- 4 labels show $20\% < d < 30\%$
- 1 label shows $d > 30\%$

Figure 7 shows that the average discrepancy related to the ranking generated by the NLP tool is 10%, the lowest value when compared with the average discrepancy of individual expert rankings. Consequently, the tool seems to be less subjective if compared with individual experts opinion mirroring the collective capability of the group to translate the quality-related natural language expressions
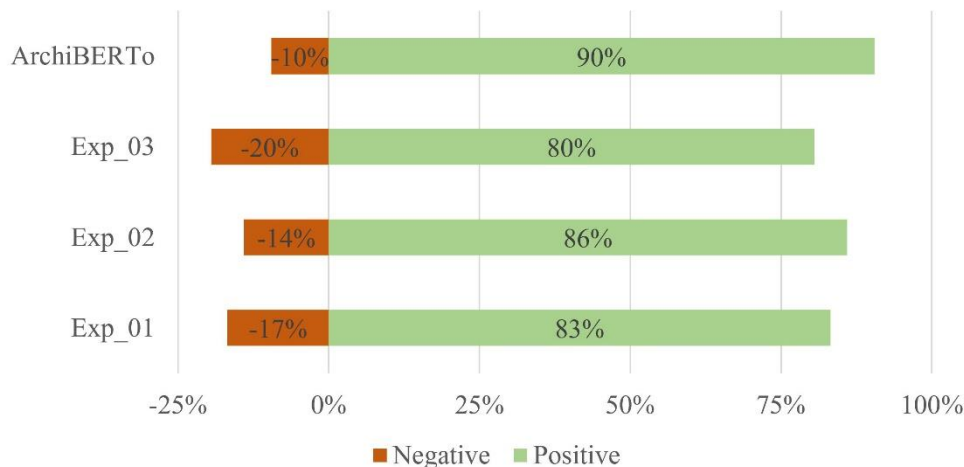

Figure 7: Single experts and NLP tool total discrepancy, DIP_01.

4.2.2. DIP number 2: Nuoro secondary school

ArchiBERTo is applied to process a DIP of a secondary school located in the municipality of Tortolì, province of Nuoro Sardinia. The objectives ranking generated by the NLP tool gets for 10 out of 21 labels/objectives the lowest discrepancy. 7 labels show a discrepancy lower than the discrepancy of the evaluation of two expert, 3 labels, C.1), E.1), and M.2), show a discrepancy lower than a single expert and for only 1 objective, G.1), the NLP tool obtains the highest discrepancy compared with the benchmark. Moreover, considering the discrepancy (d) of the ranking provided by the NLP tool:

- 13 labels show $d < 10\%$
- 7 labels show $10\% < d < 20\%$

- 3 labels show $20\% < d < 30\%$
- 1 label shows $d > 30\%$

Figure 8 shows that the average discrepancy related to the ranking generated by the NLP tool is 11%, the lowest value when compared with the average discrepancy of individual expert rankings. The results of the second DIP processed confirm the lower subjectivity of the NLP tool and its capability of mirroring the collective knowledge of the group of experts in the objectives ranking task.
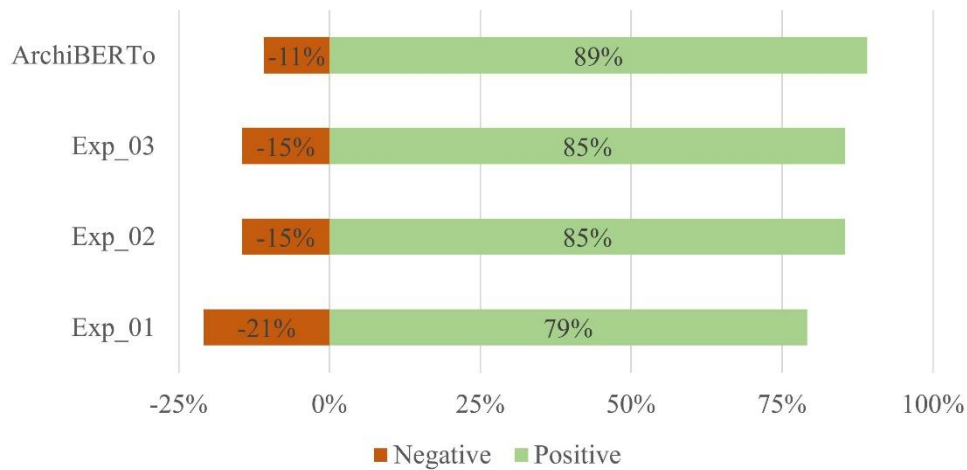

Figure 8: Single experts and NLP tool total discrepancy, DIP_02.

4.3. ArchiBERTo evaluation: customization degree
4.3.1. DIP number 3-4-5

The capability of the NLP system to generate rankings customized on different DIPs is evaluated measuring the variation of the rankings generate processing three different DIPs compared with a fixed ranking produced by the three experts collectively. The variation obtained for the DIPs is shown in Figure 9.
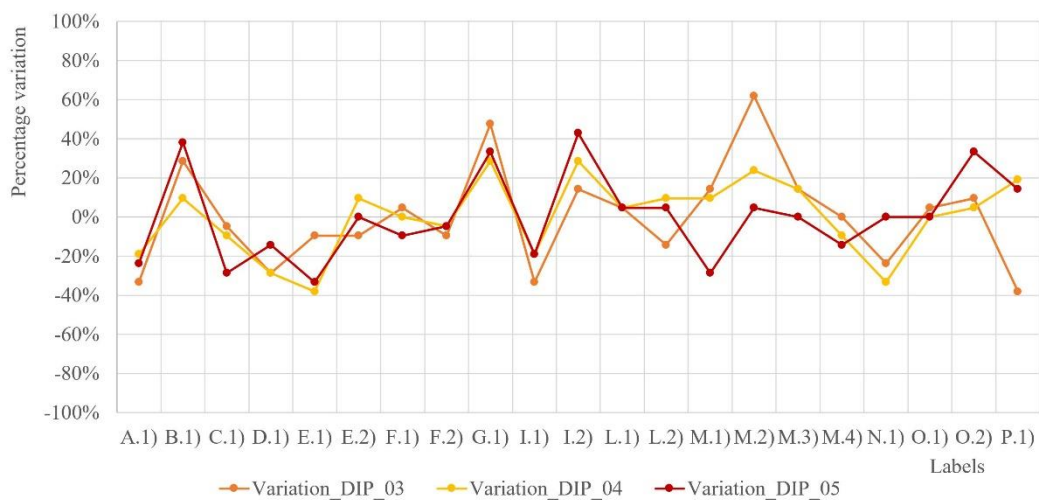

Figure 9: DIPs ranking variation comparison with a fixed ranking.

The DIP analyzed and translated by the group of experts into the fixed ranking used to compare the ArchiBERTo outputs related to the three different DIPs concerns the design and construction of a secondary school building. The DIP_03 concerns the design of a secondary school as well as for the DIP_05, the DIP_03 is about the construction of a new primary school.

The NLP tool seems to show a good customization degree with an average variation of 20, 15, and 17% for the three DIPs analyzed respectively.

5. Conclusion

The study aims to demonstrate the capability of an ad-hoc developed NLP tool (ArchiBERTo) to automatically process and translate quality objectives expressed in DIPs of school building projects into a list of hierarchized objectives to support the evaluation process of the design proposals in the Italian design and construction context of Progetto Iscol@. The use of ArchiBERTo aims to establish a consensus between the involved actors (i.e., public appointing party, design teams and, the evaluation committee) regarding the relative hierarchy of quality needs and objectives providing a list of ranked objectives and criteria, enhancing the effective communication during the pre-design phase.

According to the results, ArchiBERTo seems to be capable of mirroring the collective capability and sensitivity of a group of experts in the architecture and construction knowledge domain in the ranking of sentences related to quality objectives, avoiding subjectivity in the interpretation of textual information. Moreover, ArchiBERTo, being the representation of a group of experts knowledge is less affected to subjectivity, outperforms the capability of a single expert to handle the complexity of analyzing several sentences contained in a DIP document.

Regarding the capability to produce customized rankings based on the content of different DIPs, ArchiBERTo shows a good degree of customization. This confirms the proper flexibility of the proposed system, which does not flatten the ranking of objectives, deviating from the use of a fixed evaluation grid as, on contrary, was the case for the evaluation of the first cycle of the call for tenders and pilot projects of the Progetto Iscol@. In fact, ArchiBERTo is able to provide a customized prioritization of objectives for the different processed DIPs, mirroring the semantic content of each documents reintroducing the proper flexibility and compliance with the different projects specific quality demands.

In the Italian public tender procedure, the prioritization ranking generated by the NLP tool can be shared along with the tender documentation to the design teams participating to the call for tenders to improve the communication and allow the designers to have full understanding of the appointing party needs and quality objectives, and to the external committee to have a support in the evaluation and comparison of the design projects.

Moreover, considering the aforementioned limitation of the BIM approach in managing unstructured information, such as the information shared in natural language in the qualitative section of a DIP document, the proposed methodology stands as an attempt to expand the digitalization of the design and construction process to the unstructured natural language data. In fact, the combined use of BIM and NLP methodologies and tools could help architects and engineers to digitally manage both aspects of the design and construction process: the alphanumeric (structured) side and the non-alphanumeric (unstructured) side of the data, the latter a key step in the digitization of the design and construction industry.

6. Acknowledgements

7. References

[1]   R. R. Senescu, J. R. Haymaker, S. Meža, M. A. Fischer, Design Process Communication Methodology: Improving the Effectiveness and Efficiency of Collaboration, Sharing, and Understanding, Journal of Architectural Engineering. 20 (2014) 1–14. doi:10.1061/(ASCE)AE.1943-5568.0000122.

[2]   N. Norouzi, M. Shabak, M. R. Bin Embi, T. H. Khan, The Architect, the Client and Effective Communication in Architectural Design Practice, in: Procedia - Social and Behavioral Sciences, Elsevier B.V., 2015, pp. 635–642. doi:10.1016/j.sbspro.2015.01.413.

[3]   H. Taleb, S. Ismail, M. H. Wahab, W. N. M. W. M. Rani, Communication management between architects and clients, in: AIP Conference Proceedings, 2017, pp. 1–6. doi:10.1063/1.5005469.

[4]   G. M. Di Giuda, M. Locatelli, E. Seghezzi, Natural Language Processing and BIM In AECO Sector: A State Of The Art, in: Proceedings of the Fifth Australasia and South-East Asia Structural Engineering and Construction Conference, ISEC Press, Christchurch, New Zealand, 2020, pp. 1–6. doi:10.14455/ISEC.2020.7(2).CON-22.

[5]   S. Sun, L. Li, Application of Deep Learning Model Based on Big Data in Semantic Sentiment Analysis, in: The 2021 International Conference on Machine Learning and Big Data Analytics for IoT Security and Privacy. SPIoT 2021. Lecture Notes on Data Engineering and Communications Technologies, Springer Science and Business Media Deutschland GmbH, Shanghai, China, 2022, pp. 590–597. doi:10.1007/978-3-030-89508-2_76.

[6]   M. Locatelli, G. Pattini, E. Seghezzi, L. C. Tagliabue, D. G. Giuseppe Martino, NLP-based system for automatic processing of quality demands in Italian public procedure: a system engineering formalization, in: 2022 European Conference on Computing in Construction, Ixia, Rhodes, Greece, 2022, pp. 1–8. doi:10.35490/EC3.2022.176.

[7]   F. Ameziane, Information system for building production management, International Journal of Production Economics. 64 (2000) 345–358. doi:10.1016/S0925-5273(99)00071-7.

[8]   A. A. Latiffi, J. Brahim, M. S. Fathi, The Development of Building Information Modeling (BIM) Definition, Applied Mechanics and Materials. 567 (2014) 625–630. doi:10.4028/www.scientific.net/AMM.567.625.

[9]   T. Young, D. Hazarika, S. Poria, E. Cambria, Recent trends in deep learning based natural language processing, IEEE Computational Intelligence Magazine. 13 (2018) 55–75. doi:10.1109/MCI.2018.2840738.

[10]   C. A. Montgomery, Linguistics and Automated Language Processing, in: International Conference on Computational Linguistics COLING, 1969, pp. 1–25.

[11]   M. Pacak, A. W. Pratt, The function of semantics in automated language processing, in: SIGIR '71: Proceedings of the 1971 International ACM SIGIR Conference on Information Storage and Retrieval, 1971, pp. 5–18. doi:10.1145/511285.511288.

[12]   J. Barnett, K. Knight, I. Mani, E. Rich, Knowledge and Natural Language Processing, Communications of the ACM. 33 (1990) 49–63. doi:10.1145/79173.79177.

[13]   A. Lenci, S. Montemagni, V. Pirelli, Testo e computer. Elementi di linguistica computazionale, 6th. ed., Carocci editore@Aulamagna, Rome, 2005.

[14]   C. Wu, X. Li, Y. Guo, J. Wang, Z. Ren, M. Wang, Z. Yang, Natural language processing for smart construction: Current status and future directions, Automation in Construction. 134 (2022). doi:10.1016/j.autcon.2021.104059.

[15]   F. U. Hassan, T. Le, Automated Requirements Identification from Construction Contract Documents Using Natural Language Processing, Journal of Legal Affairs and Dispute Resolution in Engineering and Construction. 12 (2020). doi:10.1061/(ASCE)LA.1943-4170.0000379.

[16]   F. ul Hassan, T. Le, Computer-assisted separation of design-build contract requirements to support subcontract drafting, Automation in Construction. 122 (2021). doi:10.1016/j.autcon.2020.103479.

[17]   T. Mahfouz, A. Kandil, S. Davlyatov, Identification of latent legal knowledge in differing site condition (DSC) litigations, Automation in Construction. 94 (2018) 104–111. doi:10.1016/j.autcon.2018.06.011.

[18]   J. Lee, J.-S. Yi, J. Son, Development of Automatic-Extraction Model of Poisonous Clauses in International Construction Contracts Using Rule-Based NLP, Journal of Computing in Civil Engineering. 33 (2019) 04019003. doi:10.1061/(ASCE)CP.1943-5487.0000807.

[19]   J. Lee, Y. Ham, J.-S. Yi, J. Son, Effective Risk Positioning through Automated Identification of Missing Contract Conditions from the Contractor's Perspective Based on FIDIC Contract Cases, Journal of Management in Engineering. 36 (2020) 1–11. doi:10.1061/(ASCE)ME.1943-5479.0000757.

[20]   R. Khalef, I. H. El-adaway, Automated Identification of Substantial Changes in Construction Projects of Airport Improvement Program: Machine Learning and Natural Language Processing Comparative Analysis, Journal of Management in Engineering. 37 (2021) 1–15. doi:10.1061/(ASCE)ME.1943-5479.0000959.

[21]   H. Fan, H. Li, Retrieving similar cases for alternative dispute resolution in construction accidents using text mining techniques, Automation in Construction. 34 (2013) 85–91. doi:10.1016/j.autcon.2012.10.014.

[22]   B. Zhong, X. Pan, P. E. D. Love, L. Ding, W. Fang, Deep Learning and network analysis: Classifying and visualizing accident narratives in construction, Automation in Construction. 113 (2020) 103089. doi:10.1016/j.autcon.2020.103089.

[23]   A. Ajayi, L. Oyedele, H. Owolabi, O. Akinade, M. Bilal, J. M. Davila Delgado, L. Akanbi, Deep Learning Models for Health and Safety Risk Prediction in Power Infrastructure Projects, Risk Analysis. 40 (2020) 2019–2039. doi:10.1111/risa.13425.

[24]   F. Zhang, H. Fleyeh, X. Wang, M. Lu, Construction site accident analysis using text mining and natural language processing techniques, Automation in Construction. 99 (2019) 238–248. doi:10.1016/j.autcon.2018.12.016.

[25]   A. J. P. Tixier, M. R. Hallowell, B. Rajagopalan, D. Bowman, Construction Safety Clash Detection: Identifying Safety Incompatibilities among Fundamental Attributes using Data Mining, Automation in Construction. 74 (2017) 39–54. doi:10.1016/j.autcon.2016.11.001.

[26]   A. J. P. Tixier, M. R. Hallowell, B. Rajagopalan, D. Bowman, Application of machine learning to construction injury prediction, Automation in Construction. 69 (2016) 102–114. doi:10.1016/j.autcon.2016.05.016.

[27]   H. Baker, M. R. Hallowell, A. J. P. Tixier, AI-based prediction of independent construction safety outcomes from universal attributes, Automation in Construction. 118 (2020) 103146. doi:10.1016/j.autcon.2020.103146.

[28]   A. J. P. Tixier, M. R. Hallowell, B. Rajagopalan, D. Bowman, Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports, Automation in Construction. 62 (2016) 45–56. doi:10.1016/j.autcon.2015.11.001.

[29]   A. Ferrari, G. Gori, B. Rosadini, I. Trotta, S. Bacherini, A. Fantechi, S. Gnesi, Detecting requirements defects with NLP patterns: an industrial experience in the railway domain, Empirical Software Engineering. 23 (2018) 3684–3733. doi:10.1007/s10664-018-9596-7.

[30]   A. Faraji, M. Rashidi, S. Perera, Text Mining Risk Assessment–Based Model to Conduct Uncertainty Analysis of the General Conditions of Contract in Housing Construction Projects: Case Study of the NSW GC21, Journal of Architectural Engineering. 27 (2021) 1–17. doi:10.1061/(asce)ae.1943-5568.0000489.

[31]   J. Lee, J.-S. Yi, Predicting Project's Uncertainty Risk in the Bidding Process by Integrating Unstructured Text Data and Structured Numerical Data Using Text Mining, Applied Sciences. 7 (2017) 1–15. doi:10.3390/app7111141.

[32]   M. Bilal, L. O. Oyedele, Big Data with deep learning for benchmarking profitability performance in project tendering, Expert Systems with Applications. 147 (2020) 1–19. doi:10.1016/j.eswa.2020.113194.

[33]   M. F. F. Siu, W. Y. J. Leung, W. M. D. Chan, A data-driven approach to identify-quantify-analyse construction risk for Hong Kong NEC projects, Journal of Civil Engineering and Management. 24 (2018) 592–606. doi:10.3846/jcem.2018.6483.

[34]   Y. Zou, A. Kiviniemi, S. W. Jones, Retrieving similar cases for construction project risk management using Natural Language Processing techniques, Automation in Construction. 80 (2017) 66–76. doi:10.1016/j.autcon.2017.04.003.

[35]   R. Zhang, N. El-Gohary, A machine learning-based method for building code requirement hierarchy extraction, in: Canadian Society for Civil Engineering Annual Conference, CSCE 2019, Laval, Canada, 2019, pp. 1–10.

[36]   J. Zhang, N. M. El-Gohary, Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking, Automation in Construction. 73 (2017) 45–57. doi:10.1016/j.autcon.2016.08.027.

[37]   M. Locatelli, E. Seghezzi, L. Pellegrini, L. C. Tagliabue, D. G. Giuseppe Martino, Exploring Natural Language Processing in Construction and Integration with Building Information Modeling: A Scientometric Analysis, Buildings. 11 (2021) 1–33. doi:10.3390/buildings11120583.

[38]   M. Locatelli, G. Pattini, L. Pellegrini, S. Meschini, D. Accardo, Fostering the consensus: a BERT-based Multi-label Text Classifier to support agreement in Public design call for tenders, 2022. To appear.

[39]   J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, ArXiv. (2019) 16. doi:10.48550/arXiv.1810.04805.

[40]   R. Venkatesan, M. J. Er, Multi-label classification method based on extreme learning machines, in: 2014 13th International Conference on Control Automation Robotics and Vision, ICARCV 2014, 2014, pp. 619–624. doi:10.1109/ICARCV.2014.7064375.

[41]   M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Information Processing and Management. 45 (2009) 427–437. doi:10.1016/j.ipm.2009.03.002.

[42]   M. L. Osborne, A Modification of Veto Logic for a Committee of Threshold Logic Units and the Use of 2-Class Classifiers for Function Estimation, Oregon State University, 1975.