DEVELOPMENT AND IMPLEMENTATION OF A PREDICTIVE MODEL OF ACCIDENT RISK ON CONSTRUCTION SITES

First A. Drd. Mădălina-Giulia BOBOCEA, *National University of Science and Technology POLITEHNICA Bucharest*
Second B. Prof. dr. ing. Oana Roxana CHIVU, *National University of Science and Technology POLITEHNICA Bucharest*
Third C. Drd. Elena- Cristina DEDIU, *National University of Science and Technology POLITEHNICA Bucharest*
Fourth D. Conf. dr. ing. Marinela MARINESCU, *National University of Science and Technology POLITEHNICA Bucharest*
Fiveth E. Conf. dr. ing. Larisa BUTU, *National University of Science and Technology POLITEHNICA Bucharest*
Sixth F. Drd. Andrei IACOB, *National University of Science and Technology POLITEHNICA Bucharest*

ABSTRACT: The article presents the development and implementation of a predictive model of accident risk on construction sites, adapted to the specific conditions of Romania. The proposed model estimates the probability of an unwanted event occurring at the team and day level, using technical, organizational, behavioral and ergonomic factors. The performance of the model is demonstrated by standard indicators (AUC-ROC, calibration curve, Brier score) and by practical implementation in the form of an operational dashboard. The contribution consists in the simultaneous integration of the determining factors in an interpretable and reproducible framework.

KEYWORDS: occupational health and safety; construction; predictive model; logistic regression; risk management.

1. INTRODUCTION

Construction site activities are highly variable and uncertain, driven by the rapid succession of operations, the interdependence of teams and external conditions that are often difficult to control. Factors such as sudden weather changes, working at height, the use of heavy equipment or poor coordination between subcontractors can exponentially increase the risk of accidents [1]. In addition, the pressure of deadlines and costs often requires the compression of work stages, which can lead to reduced attention to occupational health and safety (OHS) requirements.

These particularities define a complex operational context, in which traditional prevention approaches, based on periodic inspections, checklists or reactive accident analyses, become insufficient. In the absence of proactive risk anticipation, preventive interventions are carried out late, and safety resources cannot be efficiently directed towards activities with real increased risk. In this context, the need for a data-based decision support tool is outlined, capable of estimating in real time the probability of an occupational accident [2], based on the specific characteristics of the teams, the environment and the organization of the construction site.

The predictive model developed in responds to this need, by integrating a coherent set of determining factors, technical, organizational, human and environmental, in an interpretable mathematical formulation. By using logistic regression, a relationship is obtained between the probability of an accident and variables such as: working at height, weather conditions, average team experience, presence of direct supervision, level of recent training, safety culture score (SSM) [3], 5S index of order and cleanliness, as well as the density of subcontractors. Each factor is associated with a β coefficient, which expresses its influence on the final probability of risk[4].

The ultimate goal of the model is twofold: on the one hand, the daily prediction of the risk of accidents for each team or activity, and on the other hand, providing an objective basis for preventive decision-making at the site management level. In this way, the model does not replace the professional judgment of the OHS managers, but complements it through a systematic and reproducible analysis of data, supporting the shift from reaction to prevention.

The concrete result of this approach is a simple software tool (dashboard) that updates the estimated probability of risk daily, allowing for prioritization of interventions and allocation of security resources where they are most needed[5].

2. MODEL CONSTRUCTION METHODOLOGY

2.1 Purpose and working hypothesis

The predictive model developed in this research has as its main purpose the quantitative estimation of the risk of accidents on construction sites, expressed as a probability with values between 0 and 1, which reflects the level of danger associated with an activity or a team at a given moment. Through this tool, the aim is to proactively identify situations with high risk potential, before they materialize into unwanted events, thus providing the site management with data-based decision-making support.

In a field characterized by high variability and operational interdependence, such as construction, risk prediction cannot be achieved solely through descriptive methods or historical statistics[6]. The proposed model brings significant added value, as it simultaneously integrates technical factors (such as working at height or weather conditions), organizational factors (presence of direct supervision, density of subcontractors), as well as behavioral and cultural factors (level of recent training, safety culture, order and cleanliness according to the 5S principles). This multidimensional approach allows for a more comprehensive understanding of the potential causes of accidents and provides a solid analytical basis for prioritizing prevention measures.

Therefore, the working hypothesis that guided the development of the model was formulated as follows:

"The likelihood of a work accident occurring on a construction site is significantly influenced by factors such as working at height, weather conditions, team experience, the presence of direct supervision, the level of safety culture, recent training and the degree of order and cleanliness (5S)."

This hypothesis reflects the belief that the risk of an accident is not a random event, but the cumulative result of a set of interdependent conditions, which can be monitored and controlled by appropriate analytical tools. Validating the hypothesis through the proposed logistic model provides a scientific basis for moving from a reactive approach, based on post-accident reporting, to a proactive one, based on the anticipation and prevention of daily risks. In this way, the predictive model becomes a strategic tool for the integrated management of occupational safety and health on construction sites in Romania.

2.2 Theoretical foundations of the model

The construction of the predictive model of the accident risk was carried out starting from the hypothesis that the probability of an undesirable event occurring on the construction site is not the result of chance, but of the simultaneous action of a set of observable, technical, organizational and behavioral factors. In this context, binary logistic regression is one of the most widely used statistical methods for estimating the probability of an event with two possible outcomes, in this case, accident (1) or non-accident (0) .

The advantage of this method is that it allows for direct interpretation of the coefficients in the form of odds ratios (Odds Ratio), providing an intuitive measure of the influence of each predictor on the final probability. Thus, a positive coefficient indicates an increase in the risk of an accident, while a negative coefficient reflects a protective effect. This property makes the logistic model suitable not only for scientific research, but also for operational implementation in an occupational safety and health (OSH) management system [7].

The logistic model also offers the possibility of robustly evaluating its performance through established indicators, such as AUC-ROC (Area Under the Curve) which measures the model's ability to differentiate between risky and safe situations and the calibration curve, which shows how close the estimated probabilities are to the actual values observed in the field. In addition, the Brier score assesses the overall consistency of the predictions, combining elements of accuracy and calibration[8].

2.3 Dataset structure and characteristics

The model was trained on a dataset constructed at the **daily and team level,** considered the unit of observation. This granularity was chosen to capture the variability of activities, environmental conditions and team composition, factors that can significantly influence accident risk. Each record reflects a unique combination of activity type, team characteristics and the operational context specific to that day.

The dependent variable is binary and was noted as follows:

A_i=1, if an work accident happened that day
A_i=0, if not

Thus, the model estimates the probability $P(Ai=1 \mid X_i)$, that is, the chance that, given a certain set of factors X_i , an occupational accident will occur.

The independent variables were selected following an analysis of the specialized literature and consultation with OSH experts in the construction field, aiming to cover all categories of risk factors: technical, organizational, psychosocial and environmental.

Table 1. List of variables used in the statistical model

Symbol	vary	Type	Description
X_1	Working at height	Binary	1 if the team performed work at a height of over 2 m
X_2	Adverse weather conditions	Binary	1 for rainy or humid weather
X_3	Average team experience	Numerical (years)	Average seniority of team members
X_4	Supervision present	Binary	1 if an OHS coordinator was present
X_5	OSH culture score	Ordinal (1–5)	Evaluation by standardized questionnaire
X_6	Recent training	Binary	1 if the team has participated in training in the last 90 days
X_7	5S Index	Numeric (0–100)	Level of order and cleanliness according to the 5S methodology
X_8	Subcontractor density	Numeric (0–1)	Subcontractor/total staff ratio per team

Each of these predictors reflects a relevant dimension of occupational risk. For example, working at height (X_1) is one of the main factors associated with serious accidents in construction, and the density of subcontractors (X_8) is correlated with fragmentation of responsibilities and lack of coherence in the application of safety measures. In contrast, the presence of supervision (X_4), a high level of OSH culture (X_5) and the maintenance of order and cleanliness (X_7) have a protective role, reducing the risk of events.

General formula of the model

The mathematical relationship describing the model is as follows:

$$P(A_i = 1|X_i) = \frac{1}{1 + e^{-(\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij})}} \quad \textbf{(1)}$$

where:

- $P(A_i = 1 \mid X_i)$) is the estimated accident probability;
- β_0 is the intercept of the model;
- β_j are the coefficients of each independent variable X_j.

The estimation of the β parameters was performed by the likelihood maximization method, using the Newton-Raphson iterative algorithm. The model was calibrated on a training data set (75%) and validated on a test set (25%), to ensure the generalizability of the results.

3. CASE STUDY: PILOT SITE – URBAN BRIDGE PROJECT

3.1. Context and description of the construction site

For the empirical validation of the developed predictive model, a pilot site representative of the real conditions of the construction industry in Romania was selected – the Urban Bridge Project, located in a metropolitan area with high traffic density and high technical complexity. The project aimed at the construction of a 220 m long road bridge, with infrastructure works, formwork and reinforcement, scaffolding installation and execution of underground and overhead electrical installations[9].

The pilot site was carried out over a period of 90 days, between April and June 2025, being chosen for the high variability of the operations and for the possibility of detailed monitoring of the teams involved. During this period, daily data were collected regarding the activities carried out, weather conditions, team composition, level of training, degree of supervision, order and cleanliness (5S index) and occupational safety culture (OSH).

The activities were organized into four main teams:

- E1 – Infrastructure, responsible for foundation and earthworks;
- E2 – Formwork, involved in casting reinforced concrete elements;
- E3 – Scaffolding, responsible for the assembly and verification of metal structures;
- E4 – Installations, with responsibilities in the installation of cables and electrical systems.

This structure allowed the model to be evaluated in real working conditions, characterized by different levels of occupational risk. The E1 and E2 teams were mainly exposed to mechanical and physical risks, while the E3 team was marked by the risk of falling from height – one of the most important predictors identified. The E4 team recorded lower risks, but influenced by the density of subcontractors and the level of technical training.

3.2. Data structure and application methodology

The model was applied to a set of 360 observations (4 shifts × 90 days), each observation corresponding to a unique combination of shift, day, and specific work conditions. Each record included eight independent variables, selected based on theoretical analysis and their relevance to the work process:

1. *Working at height* (H) – binary indicator (1/0), associated with scaffolding and formwork work;
2. *Weather conditions* (W) – 1 for wet weather, 0 for dry conditions;
3. *Team experience* (E) – average seniority of team members in years;
4. *Supervision present* (S) – 1 if an OH&S coordinator was present in the field;
5. *Security culture* (C) – score on a scale of 1–5;
6. *Recent Training* (T) – 1 if the team has completed training in the last 90 days;
7. *5S Index (Q)* – score between 0 and 100;
8. *Subcontractor density (D)* – the ratio between the number of subcontractors and the total active personnel.

For each observation, the binary dependent variable A_i was defined as follows:

A_i=1, if an work accident happened that day
A_i=0, if not

The logistic model used was of the form:

$$P(A_i = 1|X_i) = \frac{1}{1 + e^{-(\beta_0 + \sum_{j=1}^{8} \beta_j X_{ij})}} \quad \textbf{(2)}$$

where the coefficients β_j express the influence of each variable on the final risk probability.

The data set was divided into 75% observations for model training and 25% for testing. Parameter estimation was performed by the maximum likelihood method, using the Newton-Raphson algorithm, and numerical standardization of variables was performed to avoid the dominance of values with different scales[10].

3.3 Results obtained

Coefficient analysis showed that working at height (β = +0.93, OR = 2.54) and wet weather conditions (β = +0.58, OR = 1.78) have a significant influence on increasing the risk of accidents. In contrast, the presence of direct supervision (β = −0.48, OR = 0.62), recent training (β = −0.44, OR = 0.65) and a high 5S score (β = −0.011 per unit) significantly reduce the probability of events.

The model's performance was evaluated through several indicators:

- AUC-ROC = 0.83, indicating good discrimination capacity between days with and without events;
- Brier score = 0.15, indicating a good calibration between the estimated and observed probabilities;
- Critical threshold ($P_{(crit)}$) = 0.27, used to classify risk into three levels: low, medium and high.

Figure 1 shows the ROC curve of the model, demonstrating a clear separation between low-risk and high-risk observations. Figure 2 shows the calibration curve, where the trend line largely overlaps the ideal line, confirming that the model provides realistic predictions.

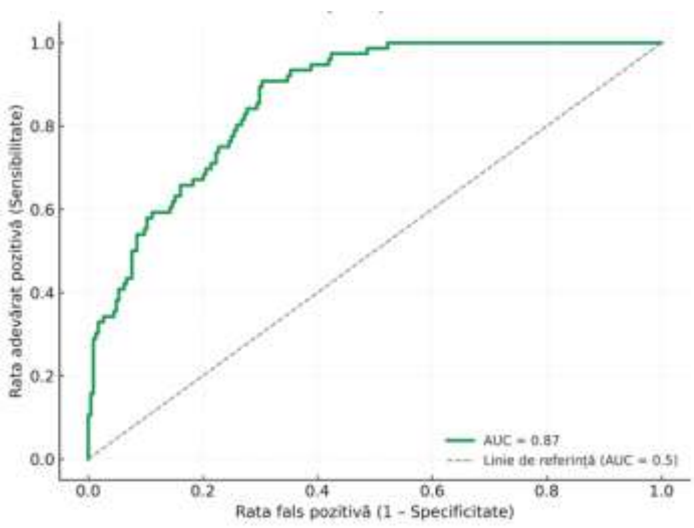

Fig. 1 – ROC curve of the model predictive

The graph illustrates the relationship between the true positive rate (sensitivity) and the false positive rate (1 - specificity) for the developed logistic model. The green line represents the model performance, with an area under the curve of 0.83, indicating a good ability to discriminate between high-risk and safe observations [11]. The diagonal gray line corresponds to the random performance (ROC = 0.5). The further the model curve deviates from this line, the higher the accuracy of the risk estimate.

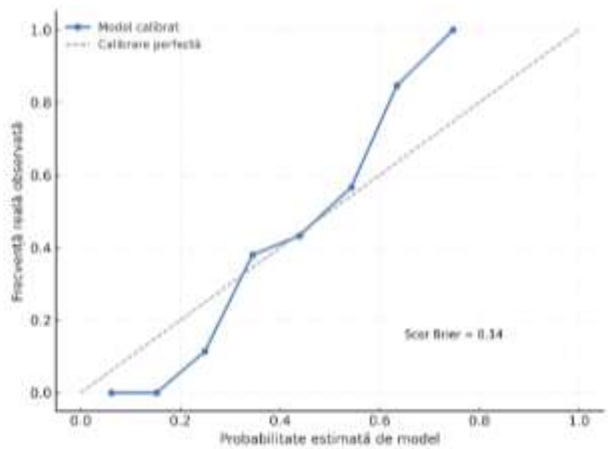

Fig. 2 – Calibration curve of the predictive model

The figure compares the model-estimated probabilities with the actual observed event frequencies. The blue line represents the calibrated model, and the dotted grey line indicates perfect calibration. The results show a good overlap between the two curves, with a Brier score of 0.15, confirming a satisfactory calibration [12]. This means that the estimated risk values (e.g. 0.30) correspond approximately to the actual accident frequency, strengthening confidence in the practical applicability of the model.

Risk dashboard was generated (Fig. 3), which summarizes the estimated probabilities for each team on a daily basis[13]. In the example from the last week of June, the E3 - Scaffolding team recorded an average risk probability of 0.34, exceeding the alert threshold, caused by the intensification of work at height and unfavorable weather conditions. In contrast, the E2 - Formwork team, with recent training and an average 5S score of 82, had an estimated risk of only 0.16.

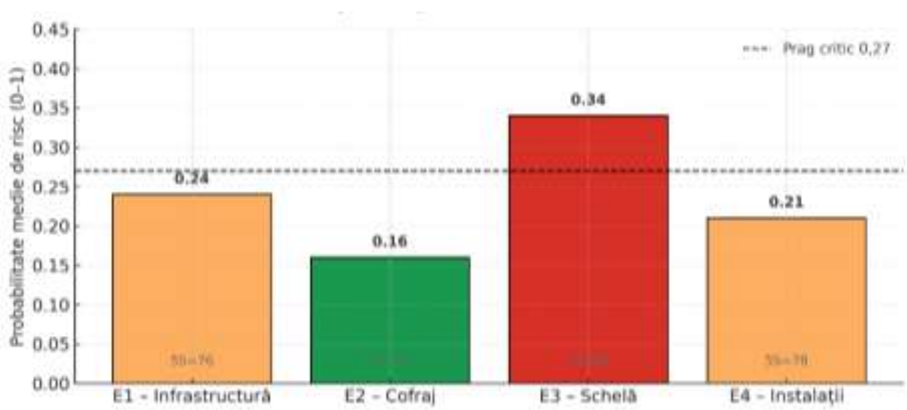

Figure 3 – Team Risk Dashboard (Pilot Site – Urban Bridge Project)

Average Probabilities for the Last 7 Days (June 2025)

The graph shows the average values of the accident risk probability for the four main teams of the pilot site, calculated for the last week of June 2025, based on the predictive logistic model. The colors used reflect the associated risk level: green (<0.20 – low risk), yellow (0.20–0.27 – moderate risk) and red (>0.27 – high risk). The black dotted line represents the critical alert threshold (P_{crit} =0.27), above which the accident probability is considered operationally significant[14].

It is observed that E3 - Scaffolding, with an average probability of 0.34, exceeds the critical threshold, confirming the increased exposure to risks specific to working at height and wet weather conditions. E1 - Infrastructure presents a moderate risk (0.24), due to mechanized works and the high density of subcontractors, while E2 - Formwork and E4 - Installations register values below the threshold (0.16 and 0.21), benefiting from better 5S scores (82 and 78) [15].

This visual representation allows for quick daily risk assessment and provides a practical tool for prioritizing inspections, scheduling training, and adjusting preventive measures at the team level[14]. The dashboard can be automatically updated based on daily site data, facilitating the transition from a reactive to a proactive approach to occupational health and safety.

3.4 Interpretation and practical implications

The results obtained support the hypothesis that the risk of accidents on the construction site is a function of the interaction between technical, organizational and behavioral factors [16], and its variation can be anticipated with a reasonable degree of accuracy through a well-calibrated logistic model. Based on the results of the pilot, the following operational conclusions were drawn:

- teams involved in work at height must be monitored as a priority on days with high humidity;
- regular training and the presence of supervisors have a clear preventive effect, reducing the estimated risk by 20–30%;
- order and cleanliness (5S) is an indirect but strong indicator of safety culture;
- Subcontractor density above 0.7 is associated with an average increase in risk of 25%, confirming the negative effects of operational fragmentation[17],[18].

Implementing the predictive dashboard in the daily HSE reporting system allows site managers to prioritize preventive actions. For example, on days when the model estimates a probability >0.27, additional measures can be automatically triggered: safety briefings, equipment checks or rescheduling of work based on weather conditions.

Case study conclusions. The application of the predictive model on the pilot site "Urban Bridge Project" demonstrated both the theoretical validity of the logistic model and its practical utility for occupational risk management. In the current context of construction digitalization, the integration of such models into OSH management systems offers a new perspective, oriented towards data-based prevention [19], not post-event reaction [20]. The applied methodology also confirms that predictive models can be used not only for retrospective statistical analysis, but also as daily decision-making tools. Through continuous recalibration, they can become part of an integrated smart safety system, adapted to the specifics of each construction site.

Relevance and limitations. The proposed model offers a balance between complexity and interpretability, being precise enough for practical application, but also transparent regarding the influence of each factor. However, it does not replace human expertise and field observations. The main limitations are related to the limited nature of the database and the need to extend testing to more construction sites for external validation.[21] However, the results obtained confirm the working hypothesis and demonstrate that the systematic integration of construction site data into a predictive logistic model can significantly support preventive decision-making, potentially reducing the number and severity of construction accidents[22].

4. CONCLUSIONS

The present research demonstrated the feasibility of developing and applying a logistic predictive model for estimating the probability of accidents on construction sites, integrating technical, organizational and behavioral factors specific to the Romanian work environment. The results obtained confirm the hypothesis that the risk of accidents is not a random event, but the cumulative result of structural and operational conditions that can be measured, monitored and managed preventively.

The proposed model, evaluated by performance indicators (ROC ≈ 0.83; Brier score ≈ 0.15), demonstrates a good capacity to discriminate and calibrate the estimated probabilities. The factors with significant influence were working at height, adverse weather conditions, lack of direct supervision, density of subcontractors and low level of order and cleanliness (5S). In contrast, the presence of recent training, a solid OH&S culture and constant supervision acted as protective factors, significantly reducing the probability of accidents.

The practical implementation in the form of an operational dashboard allows for daily risk assessment, prioritization of inspections and efficient allocation of resources for prevention. In this way, the model supports the transition from a reactive to a proactive approach, based on objective data and evidence. At the same time, the research emphasizes the importance of a safety-oriented organizational culture, in which data collection and analysis become strategic management tools, not just administrative obligations. In the future, the model can be expanded by integrating data from modern technologies (IoT sensors, video monitoring, virtual reality tools), thus strengthening the digital ecosystem of construction safety and contributing to the real reduction of work accidents.

REFERENCES

[1]. Ștefãniță PLUTEANU, Gabriela Nicoleta TANASIEV, Roxana PĂTRAȘCU, Vladimir TANASIEV, "Regulatory and economic impacts of BIM implementation in the construction industry ", U.P.B. Sci. Bull., Series C, Vol. 87, Iss. 2, 2025 ISSN 2286 – 3540

[2]. A. Moise, O. R. Chivu, M. Gheorghe, C. Borda, A. Iacob, "Study regarding the use of banks support systems in the field of construction," Annals of the „Constantin Brancusi" University of Târgu Jiu, Engineering Series, no. 4, pp. 78–82, 2024.

[3]. Ali S-Al-Ghamdi," Using logistic regression to estimate the influence of accident factors on accident severity", December 2002, Accident Analysis & Prevention 34(6):729-41

[4]. Dorinela COSTESCU, Mircea ROŞCA, Ştefan BURCIU 3, Florin RUSCĂ, “On accident prediction functions for urban road intersections”, U.P.B. Sci. Bull., Series D, Vol. 78, Iss. 3, 2016 ISSN 1454-2358.
[5]. Irinel Daniela IACOB, “Dam’s risk management”, U.P.B. Sci. Bull., Series D, Vol. 75, Iss. 3, 2013 ISSN 1454-2358.
[6]. Choi, J., Gu, B., Chin, S., Lee, JS (2020). Machine learning predictive model based on national data for fatal accidents of construction workers. Automation in Construction, 110, 102974.
[7]. Tixier, AJ-P., Hallowell, MR, Rajagopalan, B., Bowman, D. (2016). Automated content analysis of injury narratives to predict and prevent construction injuries. Safety Science, 86, 58–68.
[8]. Mostofi, F., Toğan, V., Ayözen, YE, Tokdemir, OB (2022). Construction safety risk model with construction accident network: A graph convolutional network approach. Sustainability, 14 (23), 15906.
[9]. Toptancı, Ş., Erginel, N., Acar, İ. (2023). Predicting the severity of occupational accidents in the construction industry using standard and regularized logistic regression models. NÖHÜ Journal of Engineering Sciences, 12 (3), 778–798.
[10]. Kines, P., et al. (2011). Nordic Safety Climate Questionnaire (NOSACQ-50): A new tool for diagnosing safety climate. International Journal of Industrial Ergonomics, 41 (6), 634–646.
[11]. Zohar, D. (2000). A group-level model of safety climate: Testing the effect of group climate on micro accidents. Journal of Applied Psychology, 85 (4), 587–596.
[12]. Reason, J. (2000). Human error: Models and management. BMJ, 320 (7237), 768–770.
[13]. Hollnagel, E. (2012). FRAM: The Functional Resonance Analysis Method—Modelling complex socio-technical systems. Ashgate .
[14]. Lingard, H., Cooke, T., Blismas, N. (2012). Do perceptions of supervisors' safety responses mediate the relationship between safety climate and safety outcomes in construction? Journal of Construction Engineering and Management, 138 (2), 234–241.
[15]. Seo, J., Han, S., Lee, S., Kim, H. (2015). Computer vision techniques for construction safety and health monitoring. Advanced Engineering Informatics, 29 (2), 239–251.
[16]. Hossain, MA, He, X. (2019). Predicting construction work zone crashes: Machine learning approach. Accident Analysis & Prevention, 133, 105263.
[17]. Hasanzadeh, S., Esmaeili, B., Dodd, MD (2017). Impact of construction task hazards on visual attention distribution and hazard recognition. Journal of Construction Engineering and Management, 143 (1), 04016074.
[18]. Sacks, R., Perlman, A., Barak, R. (2013). Construction safety training using immersive virtual reality. Construction Management and Economics, 31 (9), 1005–1017.
[19]. Zou, PXW, Sun, A. (2013). Management of safety risks in

construction: Review and future research. Safety Science, 57, 11–23.
[20]. Liu, Y., Liu, J., Zhang, L. (2024). Early warning for construction accidents using neural networks. Engineering Reports, 6 (2), e12654.
[21]. Iacob.A.St, Gheorghe M., CIRTINA D., Marinescu M.N., Radu C., "The impact of digitalisation on workers' safety and health" , TECHNICAL UNIVERSITY OF CLUJ-NAPOCA, ACTA TECHNICA NAPOCENSIS, Series: Applied Mathematics, Mechanics, and Engineering, Vol. XX, Issue xx, Month, 20xx
[22]. Silviu Codescu, Daniela Popescu, Valeriu Panaitescu, "Computational simulation of road tunnel fire protection, by using sprinklers", U.P.B. Sci. Bull., Series D, Vol. 77, Iss. 2, 2015 ISSN 1454-2358