https://doi.org/10.31449/inf.v49i11.9783 Informatica **49** (2025) 401–420 **401**

A Hybrid Mamdani Fuzzy Inference System and Generalized Regression Neural Network for Cost and Time Overrun Prediction in Expressway Construction Projects

Tong Yao[1] , Xiao Luo [2*]
[1]School of economics and management, Sichuan Tourism University, Sichuan, 610100,China
[2]Overseas Education College, Chengdu University, Chengdu,Sichuan,610106, China
E-mail: love_luoxiao1989@126.com
[*]Corresponding author

Keywords: cost and time overrun, mamdani fuzzy inference system, generalized regression neural network model, expressway construction and risk factors

Received: June 18, 2025

In construction projects like expressways, risks and uncertainties are unavoidable and have the potential to significantly alter the expected outcome, which would be detrimental to the project's success. Risk is one of the main causes of productivity and efficiency losses in the construction sector that results in project demise, disputes, costs, and time overruns. The failure to complete the construction project within the stipulated time and estimated cost due to various risk factors is a major problem nowadays. This work analyzes the risk factors resulting in cost and time overrun in expressways construction projects using the Mamdani Fuzzy Inference System and Generalized Regression Neural Network(H-MFIS-GRN2) hybridization. Initially, the MFIS is used to find and measure potential dangers in the building process by analysing expert-made fuzzy rules. Defuzzification of the outputs yields clear risk severity values, which are further weighted with the use of mean scores and the Relative Importance Index (RII). Learning to anticipate budget and schedule overruns, MFIS doesn't provide the final product *risk variables as normalized and prioritized inputs. The GRN2 model trains the observed risk factors with a pattern using the Gaussian activation function in a single pass. The model was validated using data from 27 areas from completed highway building projects. The trials show that the H-MFIS-GRN2 model outperforms baseline models. These baseline models are HRF-GA, H-AHP-ANN, and F-MRA. The H-MFIS-GRN2 model has 92.5% accuracy and 5.3% MAPE. Comparison analysis has increased forecast accuracy and interpretability, helping prioritize and minimize key risk variables. Fuzzy logic and neural networks can be used together to detect early risk in major infrastructure projects due to their strengths in uncertainty and learning.*

Povzetek: V članku je opisan sistem za napovedovanje prekoračitev stroškov in časa pri gradnji avtocest. Hibridni algoritem H-MFIS-GRN2 združuje Mamdanijev sistem zamegljenega sklepanja (MFIS) in regresijsko nevronsko mrežo (GRN2) za obravnavo tveganj in negotovosti.

1 Introduction

The planning, engineering, and construction of high-speed roadways specifically intended for quick and effective transit is called expressway construction. Expressways, sometimes called motorways or highways, are built to manage heavy traffic, enabling efficient vehicle movement and uninterrupted travel. The construction of expressways ranks with the highest frequency of mishaps, fatalities, delays, and cost overruns, largely because of unmanaged risks. With grade partitions at key intersections, an expressway is a segment of an urban highway with variable restrictions on entry. The term "overrun risk factors" refers to the variables that may cause schedule and expense overruns, resulting in delays and greater expenditures. These variables may change according to the project, the region, and other particulars. Risk is characterized as an unpredictable, unforeseen event or circumstance that could have an unanticipated impact on a minimum of one project objective like schedule, expense, quality performance, etc. The risk might come from the bidding process, weather changes, job site effectiveness, political circumstances, market rivalry, etc., in constructing expressway fields. An efficient risk management system can reduce its negative effects on project objectives. In expressway construction projects, overrunning the risk factors and uncertainties are unavoidable and has the potential to significantly alter the planned result, which would be detrimental to the project's success.

Understanding the root of any uncertainty in managing expressway construction projects demands going above a simple inquiry into the expenses and schedule. Key aspects like "the imprecision, ambiguity, and uncertainty of the risk variables are essential for an expressway side construction to properly cope with a contractor's risks to the project using Fuzzy Set Theory (FST). Several risk factors impacting construction project's time and cost overrun been found utilizing various statistical and computational techniques. Fuzzy logic is one such concept that can be applied. There are many unknown risk factors involved. In the case of the construction sector, it has the potential to produce reliable outcomes. Fuzzy techniques have been widely embraced as hybridized tactics for construction project risk analysis because they are incredibly effective at quantifying the risks experienced in complicated construction endeavors. The usage of neural networks is justified by their adaptability and propensity to anticipate and categorize all types of data more accurately than any other type of classifier. Hence, various neural network approaches are analyzed for predicting overrunning risk factors in expressway construction projects.

The main contributions of the article include

1. A new hybrid model (H-MFIS-GRN2) combines Mamdani-type Fuzzy Inference Systems (MFIS) with Generalized Regression Neural Networks to anticipate expressway building project cost and time overruns.
2. The model overcomes fuzzy-only models' static constraint and neural networks' black-box character by combining data-driven learning with expert-driven fuzzy rule logic, improving interpretability and prediction accuracy.
3. The authors propose a triangular fuzzy membership function, language-variable, and expert-review–derived Relative Importance Index (RII) risk prioritization method.
4. Single-pass training and Gaussian kernel activation help the GRN2 model beat baseline models like F-MRA, H-AHP-ANN, and HRF-GA with 92.5% prediction accuracy and 5.3% MAPE.
5. Validated on 27 Indian expressway projects, the model provides operational, financial, regulatory, investment, and climatic risk information. These parameters enable practical risk minimization.

The automation, productivity, and dependability of the construction sector are significantly increasing, and the sector is changing throughout the whole project life cycle, including the planning of projects, development, operation, and servicing [1]. The construction business projects, which greatly boost a nation's gross domestic product, include the transportation sector as a crucial component. Highways, primary intercity roads, toll roads, and other significant roadways with bridges, expressways, ducts, and tunnels are all included in road networks [2]. A very challenging and complex process is to analyze the time, costs, and risks that accompany a construction project while accounting for the specifics of all investments and the differing nature of its completion situations; hence fuzzy logic is applied to solve the above issues and imprecise informationin [3]. Tiruneh et al. [4] studied system training, assessment, and decision-making with neuro-fuzzy hybrid systems utilized in the predictive modeling of construction project challenges and concentrated information about the model's precision and accessibility performance. The applied neural network model was used to predict cost coated at highway construction projects' bidding stages [5]. Correlation analysis is used to identify the project attributes associated with a cost for lowering a bid, a technical, budget, etc. This method affects the dependent variables due to its limited number of factors(four) for analysis. Gondia et al. [6] analyzed and predicted the project's postponement risks in construction using Decision Tree and Naive Bayes models utilizing objective data from prior projects' time and delay influencing factors. Alawad et al. [7] proposed a computer vision with a Convolutional neural network model for safety risk assessment in railways to avoid high-risk possibilities. However, it does not focus on factors of risk overrun. To concentrate on risk factors like cost, quality, delay, and high economic risk, Andrić et al. [8] applied a combined approach of probabilistic model, fuzzy logic, and matrices with sensitivity analysis for belt and road initiative projects. The above-said method failed to concentrate on the dynamic nature of risk throughout the project. For implementing the dynamic nature of risk categories related to public and private partnership projects [9], an applied fuzzy interpretative structural model was used to create a link between the risks. Although fuzzy theory lessens the subjective aspect of expert judgment, it also makes processing information more time-consuming and expensive, necessitating more funding for modeling and analysis that affects the project's success. Hence, Isah and Kim [10] proposed a study for risk evaluation in expressway projects and incorporated an expert choice into the deep neural network to forecast the optimal cost and project allocation performance for successful outcomes. Risk assessment now includes a new source of uncertainty and unpredictability due to the expert judgment's subjectivity. Afzal [11] indicated that to more accurately represent complexity-based risk relationships under uncertainty, a mixed approach combining fuzzy logic and an enhanced form of the Bayesian belief network (BBN) may be used in the cost-risk analysis. This study exclusively addresses the individual risk evaluation methodologies used in expressway construction management to address the issue of cost overruns.

To avoid single cost factor estimation, Petroutsatou et al. [12] introduced a probabilistic approach for calculating the life-cycle expenses of road tunnels. Subsequently, tries to comprehensively capture their innate ambiguity, facilitating more trustworthy making choices at the start of a project. Once the input data are stored, a variant of the ANN model may generalize from them and learn the risk overruns in a single data pass [13]. Due to the local minimums of the variance requirement, the neural network method can converge to good solutions and is relatively easy to simulate the relationship between the risk overrun factors in construction projects. Hence, Elbashbishy et al. [14] created a set of genetic algorithms (GA2) and ANN models to evaluate the effect of construction project risks on cost overruns rather than a single-point calculation of risk factors. Even though the existing works perform well related to the analysis of overrun risk factors, there is room for advancement in each work mentioned with its research gaps that need to be solved in the proposed scheme. Inspired by the existing works, to solve the limitations of single risk factor analysis of cost, static nature of risk analysis, solely based on expert choice, and region-based analysis, the proposed model handles the dynamic nature of various risk factors based on different dimensions in all areas starting from the initial stage of the project.

The study's main objective is i) to create a model that can accurately assess the intricate relationships between numerous project risk components and their impact on cost and time overruns using the relative importance index and mean score method in the MFIS model. ii) to develop a predictive model based on project risk parameters and historical data that can precisely predict future cost and time variations using the Gaussian kernel activation function from the project plan using GRN2. iii) The results are obtained by calculating accuracy, MAPE, and influencing factors-based cost and time overrun factors.

Although the H-MFIS-GRN2 framework suggests a new way to combine neural network and fuzzy logic techniques, the study is still led by well-defined goals and testable hypotheses. The main goal of this research is to create and assess a hybrid risk prediction model for expressway building projects that uses data-driven learning (GRN2) and expert-driven fuzzy reasoning (MFIS) to anticipate when and how much money will go over budget. As a result, we arrive at the following important research topics:

(RQ1:) Is it possible for the hybrid MFIS-GRN2 model to forecast time and cost overruns based on risk variables better than current models like F-MRA, H-AHP-ANN, and HRF-GA?

(RQ2): When working with subjective and uncertain construction data, can integrating fuzzy rule-based reasoning with GRN2 increase interpretability and generalization? (RQ3): What is the efficacy of MFIS in dealing with complex and unclear risk indicators and converting them into weighted inputs for predictive modeling?

In line with this, the study postulates that

(H1): When tested on actual project data in benchmarking trials, the H-MFIS-GRN2 model will attain a prediction accuracy more than 90% and a MAPE less than 6%.

(H2) By maintaining rule-based decision traceability via fuzzy inference, the model will offer increased interpretability compared to black-box machine learning methods. At the outset, the MFIS part of the goal is to capture and quantify expert knowledge using fuzzy rules and trapezoidal membership functions so that uncertain risk dimensions (such as operational, financial, and climatic risks) may be effectively categorized. The second stage is completed by GRN2, which uses Gaussian kernel-based regression to generalize the learned patterns from previous project data. It does this by passing the defuzzified outputs, which have been weighted using the Relative Importance Index (RII). All of these parts work together to help project managers make better decisions by balancing the need for explanations with the need for accurate predictions. To make sure the methodology is in line with the study's goals, these hypotheses and objectives also serve as a foundation for comparing the hybrid model's performance to previous efforts.

The discussion of the remaining work is structured in the following way: Sect II discusses the existing research articles relevant to construction project overrun risk factors. Sect. III implements the detailed procedure of a hybrid combination of FMIS-GRN2. Section IV discusses the experimental analysis, including a comparative study analysis of various metrics. Sect. V summarizes the overall work procedure and contributions related to the research idea with the addition of future enhancement.

2 Literature background

Different fuzzy combinations have tackled individual uncertainty and unpredictability within construction project management situations. Risk factors were identified and classified through this literature review.

Chattapadhyay et al. [15] applied Genetic Algorithm with K-means clustering (GA+Kmeans), with the Euclidean distance method and silhouette coefficient for centroid optimization, to identify high-risk factors correlated with sub-risk categories. Applied statistics method for analyzing the most important features and achieved desired performance through cost-effectiveness, prompt project completion, high quality, and improved project scope. However, this research had the restriction of effectively acquiring inputs from experts since it was frequently difficult to persuade the participants regarding the relevance of their inputs, involving input from humans as expert opinion and discussions.

Yaseen et al. [16] created the Hybrid Random Forest classifier with Genetic Algorithm optimization (HRF-GA) to predict construction project challenges especially delay problems. The algorithm supports uncertain, complex, and dynamic environmental tasks and predicts the project performance based on these risk factors collected from Iraq. HRF-GA achieved 92%, 87.2%, and 0.833% of the measured accuracy, kappa, and classification error while identifying the delay of the construction project. It also identified precision and recall. Considering how these outside influences, like environmental changes, may affect the delay prediction model's accuracy is crucial.

El-Kholy [17] investigated the most effective models for forecasting delays and overruns in cost percentages for expressway projects using different ANNs-based algorithms: Principal Component Analysis (PCA), modular N2, Radial Basis Function (RBF) generalized regression model/ probabilistic N2, and time-lag recurrent network. The research results produce a MAPE of 25% of cost overrun. The research limitation is that it is not applicable globally due to the construction's rapid growth environment and requires frequent updates regarding factors that impact cost and delay overrun.

Hung [18] applied the Artificial Neural Network (ANN) method for risk evaluation in construction projects to help contractors at the initial stage. The failure mode and effect analysis approach is used for assessing risk factors, while the ANN technique measures the effect of risk on contractor profit. The result shows that the minimum error rate during the training period is 0.004, and for testing is 0.035. Although ANNs are good at capturing complicated correlations and patterns in the data, it can be difficult to comprehend the reasoning underlying the predictions made by the model and pinpoint the precise elements that go into the risk assessment.

Lin et al. [19] employed a hybridized Analytic Hierarchy Process (AHP), and ANN(H-AHP-ANN) was employed to develop a framework for predicting the 19 important risk factors of the construction quality of Taiwan projects. The weightage of significant risky elements to confirm their impact on construction was determined using AHP. The ANN was used to identify the characteristics of significant risk indicators and estimate the caliber of a construction project. The suggested approach produced an accuracy of 0.852% with ANN and construction audit data to forecast outcomes related to project quality; nevertheless, a weakness of the study is that the choice of risk indicators relied on experts or authorities.

Sharma et al. [20] analyzed the level of risk that overrunning of expense elements has been estimated using a new fuzzy-based approach, which considers the problems' ambiguity, uncertainty, and subjective nature. The degree of severity and likelihood ratio determines the risky elements contributing to expense overrun. An innovative measure for overruns in cost elements indicates a particular factor's risk magnitude, known as the fuzzy Index for cost overrun. The research drawbacks are there may have been more respondents picked for the study. There were also very few professionals to provide their opinions based on the sector's size.

Ashtari et al. [21] employed a Bayesian Network (BN) classifier model to forecast cost overrun and evaluate risks by considering potential interactions between predictors. The current analysis determined that inflation, a rise in the cost of resources, and a lack of experience and knowledge among the workforce were the three most important threats with risk as an input factor. According to the findings, the 18 BN models had an average accuracy of predicting 79.1% with 10fold cross-validation of training and evaluation methods. Before using the model in other contexts, one should carefully assess its relevance to various construction projects, locales, or periods.

Yun et al. [22] gathered socio-geospatial features from various data sets and built a Random Forest (RF) model to find their relationships with cost overrun. The developed models identify extremely important characteristics that affect overrun of cost, such as the original sums, unique time frame, management regions, total number of lane sets, commuting sequences, technological terrain, and temperatures on average, demonstrating that social and economic circumstances have a significant impact on real project expenditures. Lack of focus on other elements, including project management techniques, contractor efficiency, and decision-making processes.

Zafar et al. [23] demonstrated the Fuzzy Synthetic Evaluation (FSE) of the construction site's time overrun risk factor. The most important factor categorizing is stakeholder influence and threats to security, in which unproductive workers and workers, inefficient scheduling and agreements, and shortages of construction materials follow. Due to the safety circumstances, the study sought the perspectives of fewer professionals and business specialists, and time overrun risk factors are applied only to a specific region; hence it is not a generalized approach.

Sharma et al. [24] employed the average medium Fuzzy influencing score and the relevant frequency index result by Multiple Regression Analysis(F-MRA) models. The time overrun-inducing aspects were categorized into risk groups, notably red, yellow, and green zone. The created model can predict the anticipated time overrun of any forthcoming construction project with a confidence level of 79.6%. The model identified the top 5 risk influencing factors for cost overrun through the survey: land usage, material shift, under traffic, lack of a proper plan, and blueprint change. Due to the changing nature of construction projects, the main causes of time delays need to be updated regularly.

Informatica has been employing fuzzy systems to facilitate infrastructure decisions. The MFIS module's fuzzy inference is congruent with Chatterjee Performance results are and Banerjee's (2023) expert system for dynamic risk assessment in infrastructure building, highlighting the requirement for adaptive rule logic in uncertain situations. Kaur & Kaur (2023) use fuzzy-AHP and TOPSIS to make multi-criteria assessments, like our MFIS-GRN2 hybrid model. This supports merging expert-driven and data-driven methodologies.

Due to its complementarity, the hybrid MFIS-GRN2 model is employed in limited data, expert knowledge, and uncertainty settings. Mamdani-type Fuzzy Inference System (MFIS) lets domain specialists embed qualitative rules using language variables, making risk assessment more apparent and explicable. This improves interpretation. This is especially beneficial for infrastructure projects since subjective judgment and risk choices must be traceable. However, data-driven Adaptive Neuro-Fuzzy Inference Systems (ANFIS) hide the taught rule structure, making them difficult to grasp. In low-data areas, ANFIS needs larger training datasets to converge. Due to its non-parametric nature, Generalized Regression Neural Networks (GRN2) with strong generalization are chosen over Genetic Algorithm-based models like HRF-GA for small to medium datasets. Unlike backpropagation networks, GRN2 learns quickly and approximates functions smoothly without weight adjustment. For sparse data cost and time overrun prediction, this is essential. By combining MFIS for fuzzy rule-based reasoning with GRN2 for data-driven regression, the hybrid framework addresses the drawbacks of opaque (GA-based) or dense data (ANFIS) models. The consequence is interpretability and forecast accuracy.

Table 1: Literature survey

Method	Model Type	Dataset / Domain	Performance (Accuracy / MAPE)	Key Features	Limitations
HRF-GA (Yaseen et al., 2020)	Hybrid Random Forest + Genetic Algorithm	Risk delay data from Iraqi construction projects	Accuracy: **92%** Kappa: 87.2%	Handles complex and uncertain delay factors	Region-specific; lacks interpretability of rule-based systems
H-AHP-ANN (Lin et al., 2022)	Analytic Hierarchy Process + Artificial Neural Network	19 quality risk factors in Taiwanese projects	Accuracy: **85.2%**	Combines expert ranking with data-driven learning	Risk indicators rely solely on expert judgment
F-MRA (Sharma et al., 2021)	Fuzzy Multiple Regression Analysis	Survey-based time overrun data	Accuracy: Not reported Confidence: **79.6%**	Zones (red/yellow/green) to prioritize time risks	Static rules; lacks adaptive learning from data
BN Classifier (Ashtari et al., 2022)	Bayesian Network	Cost-risk interdependencies in various construction sites	Accuracy: **79.1%** (10-fold CV)	Captures probabilistic relationships between risks	Model calibration is complex and dataset-specific
ANN Models (El-Kholy, 2021)	PCA, RBF, Time-lagged ANN	Expressway delay and cost %	MAPE: **25%**	Evaluates time-lag impact in project delays	Frequent updates needed; not globally generalizable
FSE (Zafar et al., 2022)	Fuzzy Synthetic Evaluation	Highway projects in conflict zones	Accuracy: Not quantified	Emphasizes stakeholder influence and security risks	Limited to regional, socio-political factors
H-MFIS-GRN2 *(proposed model)*	**Hybrid MFIS + GRN2**	27 Indian expressway projects	**Accuracy: 92.5%, MAPE: 5.3%**	Combines expert rules and learning; handles dynamic risk profiles	Limited sample size; more projects needed for generalization

3 Proposed scheme

Risks associated with construction projects include environmental effects, financial overruns, delays, and safety risks. Due to their distinctive features, such as independence, diversity, and uncertainty, construction projects are complicated and fraught with risk. Construction projects interact with the surrounding environment rather than running independently. If all of the data is unknown, the issue is uncertain. Because most of the risks ssociated with building projects are stochastic and unclear, developing an effective algorithm and decision-making process is impossible. The success of a project depends on early cost estimation and cost management during the construction phases. Unfortunately, cost overruns in expressway construction projects are widespread and occur under various economic and regulatory circumstances, which frequently harm achieving project objectives.

GRN2 avoids local minima traps and converges faster than iterative backpropagation-based ANN models. It excels in infrastructure risk assessments and other sectors with minimal training data and noisy inputs. GRN2 is computationally cheap, generalizable, and accurate, unlike GA-based models like HRF-GA, which need extensive tuning.

Figure 1 analyzes the possible risk factors from various expressway construction projects. The collected input variables are passed to the MFIS model to identify risks and uncertainty in the construction investment process and those related to construction initiatives timing and cost problems. An accurate estimate of potential risks guarantees the accomplishment of the project's objectives. The analysis can include linguistic factors and expert knowledge to produce a rule-based system using an MFIS. Those experts were chosen based on their professional backgrounds, and administration positions in the construction industry may involve assistant engineers and deputy engineers. This system can include a variety of input factors. Both fuzzy sets and functions for membership can be used to define these variables. The fuzzy inference system then processes the input variables. It establishes the membership level for every result factor, specifically the cost and time overrun, using fuzzy logic operations like fuzzy rules and implications. The output variables have distinct fuzzy sets and functions of membership defined for them. A knowledge base is a body of information compiled through human experience and used to simulate how the inputs and results of a process related to one another. Rules are used to communicate knowledge, and using linguistic variables is the most prevalent rule framework in MFIS.

MFIS converts expert language evaluations into numerical risk weights that the GRN2 model may use as structured input features to anticipate numerical cost and time overruns.

A particular kind of neural network that works well for tasks involving regression analysis is the GRN2. It makes use of a network architecture known as radial basis function (RBF), which is able to generalize well by learning from the correlations between input and output data. The non-linear correlations between the input factors, which include project scope, operation detail, technical knowledge, weather conditions, etc., and the related cost and time overruns can be captured efficiently by the GRN2.

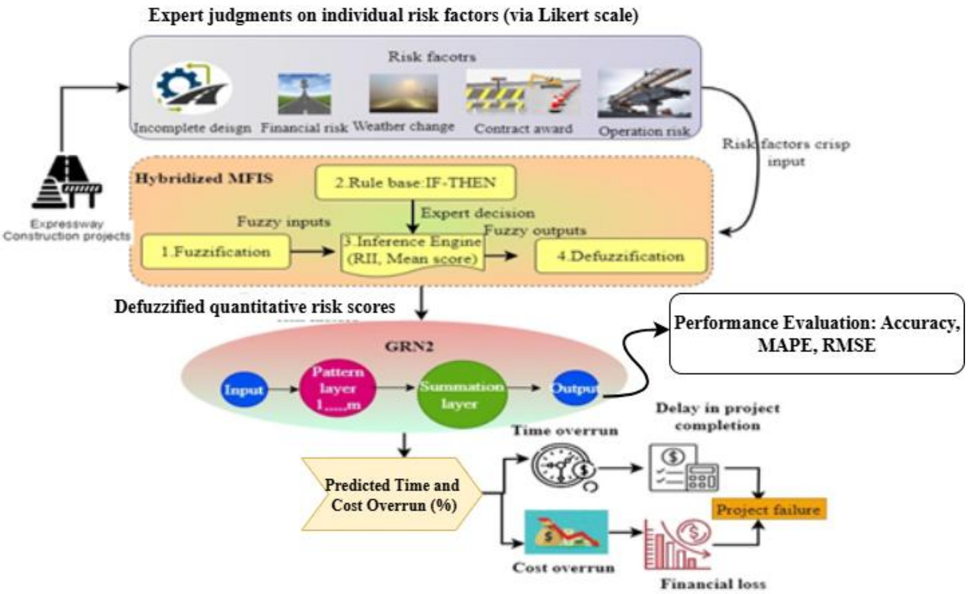

Figure 1: Systematic approach of H-MFIS-GRN2 Algorithm

With this GRN2 model, the four layers are employed to predict the cost and time overrun due to the influencing of risk factors applied in the training of GRN2. If a cost overrun occurs, it leads to financial loss that affects the economic condition, and similarly, the time overrun delays the completion of the project schedule. If both these conditions occur in construction projects, it leads to project failure. This emphasizes the importance of finding suitable risk reduction techniques in the future.The MFIS module utilizes relative significance indexing, mean score aggregation, and fuzzy inference based on expert-defined rules to provide a collection of defuzzified, weighted risk factors, as shown in Figure 1. These outputs are structured input characteristics used to train the GRN2 prediction model; they are prioritized representations of construction hazards.

Figure 1 clearly labels the GRN2 output node as "Prediction output (accuracy, MAPE)" for simplicity. The GRN2 model does, however, directly forecast percentages of both time and expense overruns. The image displays the post-prediction assessment metrics that contrast the model's predictions with the dataset's actual overruns. In this context, the terms "accuracy" and "MAPE" are utilized.

Despite its broad definitions of "risk overrun" and "project success," the MFIS is not utilized alone to forecast project results. Instead, a semantic pre-processing layer rates risk factor severity using expert-based fuzzy inference. The Mean Score and Relative Importance Index refine these numerical risk severity values from these language assessments after defuzzification. Final time and cost overrun predictions are made by the GRN2 model using a feature vector constructed with defuzzified weighted risk variables. MFIS doesn't provide the final product, but it gives GRN2 standardized and interpretable risk representations.

3.1 Mamdani type of fuzzy inference system (MFIS)

The MFIS model's functionality is based on three consecutively running blocks: fuzzification, inference engine, and defuzzification.

Step 1 shows that the MFIS can handle discrete risk factor inputs such scope quality, climate effect, technical competency, etc. Inserting input variables and linguistic terms into a normalized [0,1] domain.

Step 2 a triangle membership functions fuzzify risk factors. It appears that the MFIS turns human risk severity evaluations into fuzzy integers, which are then defuzzified. Many fuzzy rules utilize expressions like "time overrun" or "project success," they better represent language judgments of risk factor severity than project results. To prioritize inputs for the GRN2 model, the MFIS solely fuzzifies, processes, and defuzzes risk factor variables, resulting in normalized severity ratings. The hybrid design isolates MFIS (rule-based semantic reasoning).

Although "time overrun" and "project success" are employed to express fuzzy rules, they truly refer to language judgments of risk factor severity rather than project outcomes. All input risk factor variables are fuzzified, analyzed, and defuzzified by MFIS. Normalized severity ratings are priority GRN2 model inputs. GRN2 and rule-based semantic reasoning (MFIS) are separated in the hybrid design.

$$\mu_{T(a)} = \begin{cases} 0, \quad a \leq x \\ \frac{a-x}{y-x}, \quad x \leq a \leq y \\ \frac{z-a}{z-y}, \quad y \leq a \leq z \\ 0, \; z \leq a \end{cases} \tag{1}$$

Where a represents the input variable, the remaining variables x, y, and z denotes the influencing conditions to which the fuzzy rules have been derived for evaluating overrun factors of construction projects.

Figure 2 shows that the membership function plotted for risk factors as input variables with a range of [0:1]. The linguistic terms of the input variables are categorized into five levels, "very low," " low," "Moderate," "high," and "very high," to reflect the influencing level. Every point in the given input space is mapped by a fuzzy membership function curve to the extent of membership that ranges from 0 to 1. The description of the fuzzy variable with the linguistic term is explained as [Very Low: above/below 0.1; Low: above 0.3; Moderate: above 0.5; High: above 0.7; Very High: above 0.9] for the risk factors. These triangular membership functions specify each input variable's membership level to various fuzzy sets, representing meaningful risk levels.

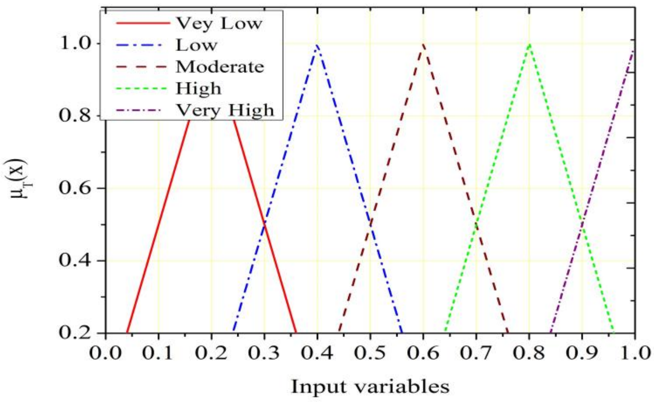

Figure 2: Membership function plots for risk factors

These $\mu_{T(a)}$ can be represented for various risk factors dimensions like operational, financial, regulation, investment, and other factors categorized for analyzing the cost and time overrun risks in the construction projects.

Step 3: Fuzzy rule-based systems use fuzzy sets and fuzzy logic to describe the humanistic knowledge about an issue and the interactions between its variables. The problem-related information that is currently known is added to the knowledge base as fuzzy IF-THEN rules. The processing component performs the inference process on the risk inputs based on the knowledge base. Managers of construction projects can receive a quantitative evaluation or prediction of the possibility of overruns in time and costs based on various influencing variables by using a fuzzy rule basis. To lessen the risk and effects of overruns, they can use this information to make wise decisions, prioritize their resources, and put the necessary mitigation measures in place.

IF x_1 is L_{t1} AND x_2 is L_{t2} OR x_3 is L_{t3} THEN Y is L_{t5}

Here the term x_1, x_2, x_3 are represented as input variables affecting risk factors, then Y is given as the output variable, then L_{t1}, L_{t2}, L_{t3} and L_{t5} are linguistic terms represented based on these variables.

Rule 1: IF the project scope is poor, THEN the risk overrun is high.

Rule 2: IF the climate is poor, THEN time overrun is high.

Rule 3: IF technical knowledge is low, AND contract design approval is poor, THEN the possibility of project success is very low.

Rule 4: IF delay in bank guarantees OR liquidity damage is high, THEN the project outcome is risk overrun is very high.

Rule 5: IF contract award is late, AND excavation works delay, THEN risk overrun is High.

Fuzzy rules that link the input parameters to output values were established in the third stage to carry out the inference. The Relative Importance Index RII_{E_c} cost and time overrun were discovered, and the relative proportion was assigned as weightage to the fuzzy rules to construct the assessment model. Give each detected risk indicator a score based on its estimated importance or possible influence on cost and time overrun. In a fuzzy membership function, Equation 2 can be used to calculate the weights from the average score of variables.

$$W_i = \frac{M_s}{\sum M_{s(\mu)}} \quad (2)$$

Where M_s represents the mean score of risk overrun factors and the $M_{s(\mu)}$ denotes the sum of all mean scores of identified risk factors in a membership function. W_i represents the weighting score for risk factor i. The percentage scores of the replies for each risk factor over the 5-point Likert scale provided by the experts are used to calculate the membership function of every risk factor.

The relative importance measure Index is a statistical method for identifying risk rank factors predicted in the expressway construction projects that impacts the project's overall success using the weighted average method. It usually ranges from 0 to 1. Project administrators can identify and concentrate on the most important risks by determining the RII_{E_c} . The priorities and issues of various stakeholders can be evaluated using the RII_{E_c}. The RII_{E_c} can direct project managers in efficiently meeting stakeholders' requirements, strengthening interaction, and enhancing final project outcomes by allocating weights to their priorities or those risk parameters, which enables efficient resource allocation and measures to mitigate risks. MFIS' fuzzy rule basis and risk factor weight allocations were based on structured feedback from six subject-matter experts. The 10-year-experienced specialists build highways and expressways. All were commercial or government infrastructure assistant engineers, project managers, or deputy engineers. Candidates excelled in building project planning, construction, and quality. The researchers used a 5-point Likert scale to assess risk factor severity. Qualitative evaluations yielded final outcomes, relative significance indices (RIIs), and fuzzy language rules. We maintained consistency using expert feedback loops since there weren't enough experts to eliminate statistical outliers. This verifies and tracks MFIS rule base data.

3.1.1 Relative importance measure index

Table 2: Likert scale type expert analysis of risk factors

No.	Causes of risk occurrences	1	2	3	4	5
1	Design changes	✗	✓	✗	✗	✗
2	Insufficient bidding method	✗	✗	✗	✓	✗
3	Failures in the design model	✗	✗	✓	✗	✗
4	Financial constraints and inadequate fund allocation from the client	✗	✗	✗	✗	✓
5	Delay in getting approvals	✓	✗	✗	✗	✗
6	Incomplete drawings	✗	✓	✗	✗	✗
7	Lack of skilled technical staff	✗	✗	✗	✓	✗
8	Deficiency of materials, equipment, and tools	✗	✓	✗	✗	✗
9	Price variation of materials	✗	✗	✓	✗	✗
10	Unfavorable climate changes	✗	✓	✗	✗	✗
11	Delay in obtaining government permits and approvals	✓	✗	✗	✗	✗

Table 1 uses symbolic indicators (✓ or X) to visualize expert selections on a 5-point Likert scale from 1 (low relevance) to 5 (high significance). Equation (2) computes the mean score (Ms) and weighted score (W) by numerically encoding replies, with each ✓ mark representing its column value (e.g., ✓ under column '4' = 4 rating). The table structure is reduced for readability because all survey responses were numerical Likert values. These numeric ratings were used to calculate the relative relevance of risk factors in MFIS fuzzification and rule prioritizing across experts.

From Table 2, construction administration experts analyze the risk factors that lead to time and cost overrun of expressways, which can be easily evaluated with the help of the mean score method. The attributes named causes of risk occurrences are evaluated on the Likert scale range from [1 to 5] where the representation of scales given by experts represented that 1 as "very low" significance for risk occurrence,2 represents the "low" importance of risk occurrence,3 represents the "moderate" possibilities of risk,4 represents "high" chances of risk, and 5 represents "very high" chances of risk overrun they choose the response option that most accurately reflects their opinion or evaluation of the specific risk factor. The pertinent risk factors that affect a particular risk assessment are identified and defined by experts. Assigning exact numerical figures to these risk factors might be difficult because they can be evaluative in form. Based on their skill, experience, and understanding, each expert assigns a rating or score to each

$$RII_{E_c} = \frac{\sum_{j=1}^{A} R_{ij}}{A \times H} \quad (3a)$$

In Equation 3a, R_{ij} denotes Likert score (1-5) assigned by expert j to factor i, A total number of experts and H is the highest possible score on the scale. The Relative Importance Index (RII) assigns a number to each risk factor. The sum of all experts' assessments for each element may be normalized by multiplying A by the maximum Likert score (H = 5). Risk factors can be prioritized when RII values are guaranteed to be [0, 1].

risk factor. Each risk factor's weight is considered after analysis when estimating the scale scores. The mean of the scores supplied by each participating expert is then used to get the mean score of risk variables. Understanding the total risk connected to each element is made easier with the help of the mean score, which offers an aggregated measurement of the expert's opinions.

To get the weighted scores, increase the scores for all risk factors by the associated weights. A higher relative relevance index (RII) suggests that the associated risk factor is more important. As illustrated in Figure 3, the number of expressway projects taken for analysis is limited to 5, with risk dimensions prioritizing the scores based on the mean score method.

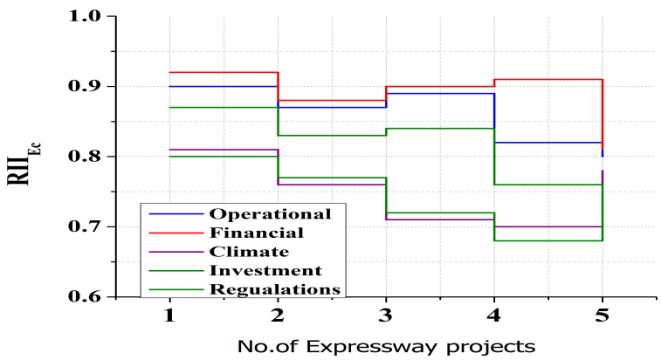

Figure 3: Analysis of projects using RII_{E_c}

Table 2 explains the ranking value given for each risk factor with its weightage assigned by experts with a clear analysis of fuzzy rules based on the assigned linguistic variables. Systems with several rules to express human reasoning and knowledge in the produced system are known as rule-based systems. Domain specialists manually built Table 2's larger risk dimensions (Operational, Financial, Climate, etc.) based on functional similarity and effect area. They then grouped Table 1's 27 granular risk indicators into five themes. The RII scores of each dimension's risk factors are summed.

Domain specialists manually built Table 2's larger risk dimensions (Operational, Financial, Climate, etc.) based on functional similarity and effect area. They then grouped Table 1's 27 granular risk indicators into five themes. The RII scores of each dimension's risk factors are summed. Specifically, for each dimension D_j, the aggregated RII is computed as in Equation 3b

$$RII_{D_j} = \frac{1}{n_j} \sum_{i=1}^{n_j} RII_i \quad (3b)$$

Where n_j is the number of risk factors belonging to dimension D_j, RII_i is the relative importance index of individual risk factor i, Qualitative reasoning and rule design were aided by MFIS linguistic mapping of these aggregated scores (e.g., Low, Medium, High influence). For the interpretation given in Table 2, Figure 7, and Figure 8, this mapping is crucial.

Table 3: RII_{E_c} based ranking

No.	Dimensions of Risk Factors	Linguistic variable	Ranking
1	Operational	High	2

2	Financial	Very High	1
3	Climate	Low	4
4	Investment	Moderate	3
5	Regulations	Very low	5

The established rule base foundation allows for the classification of cost and time/delay overrun risk into the following levels: "very low," " low," "Moderate," "high," and "very high." Various fuzzy rules were framed in the inference block to analyze the risk overrun category levels(Table 3).

Step 4: The aggregated model collected from fuzzy rules is considered fuzzy output defuzzified by the model in the fourth stage to provide crisp output values. The cost overrun percentage was calculated in the range of [0:100].

The resultant variable represents the likelihood that the costs and time of a certain construction project unit will end up being exceeded. These defuzzified risk variables are given as input to the next GRN2 prediction model to measure construction projects' cost and risk overruns accurately.This study's fuzzy rule foundation relied on domain experts' manual input, not optimization or automated rule extraction. No post-generation rule pruning or weight learning was done. The rules were constructed using language characteristics and expert consensus to assure semantic interpretability and domain knowledge alignment. The method improves transparency but isn't suitable for larger or more diversified datasets. Evolutionary algorithms and subtractive clustering may be helpful data-driven fuzzy rule learning and pruning strategies for model refinement. These approaches may dynamically optimize the rule base, reduce redundancy, and preserve interpretability.

3.2 Generalized regression neural network training model

The obtained defuzzified risk values are used for the next purpose, mainly for training the neural network model to recognize the cost and time overrun of an expressway construction project.

A generalized regression neural network (GRN2) is frequently utilized for feature estimation. It is a type of one-pass feed-forward type ANN.GRN2 represents a better method for creating neural networks based on nonparametric analysis. The GRNN uses neural networks to approximate or estimate functions and predicts the output from input data. Every training sample is supposed to represent an average corresponding to a radial base neuron. It has a specific linear layer and a radial basis layer, whereas the accuracy and speed of the GRN2 training process are advantages. GRN2 has four processing units: Input, hidden/pattern, summation /add, and output neurons. The reason for choosing this neural network model is it can handle noisy information in the input and requires only a small amount of data for training purposes.Regression analysis is used to predict time and cost overruns in this study. Comparisons to baseline estimates show ongoing percentage overruns. GRN2 converges rapidly, performs well with nonlinear regression, and approximates complex input-output mappings without repetitive backpropagation. The model converts the MFIS-derived risk factor vector to % overrun values for more accurate time and cost deviation estimates.

3.2.1 GRN2 training

Considering the fuzzy-encoded input information and the related risk overrun factors, train the GRNN portion of the Fuzzy GRNN. Analyzing the initial training data, the GRNN discovers the relationships and trends among the input factors and the cost and schedule overrun probability. While training, it memorizes every pattern of risk overrun factors in a single time pass; hence, backpropagation is unnecessary.

Table 4: Parameters of developed Grn2 model

Setting up of parameters in the GRNN model	
Input layers	i1, i2, i3, i4.
Activation function	Gaussian kernel
Type of analysis	Regression

Smoothing factor	Σ
No. of neurons in the model	21 in hidden(pattern) layer
Learning cycles	1000

From Table 4 above, the initial parameter settings of the GRN2 model are assigned, and the internal behavior of GRN2 with the obtained fuzzy rule-based output values is trained well to easily predict the various dimensional behavior of construction projects.

i) The input layer of the GRNN is the scaled representation of the fuzzy risk overrun factors. A defuzzified crisp set of a certain risk overrun factor corresponds to each input neuron in the source layer. Then it forwards the feature variables to the hidden layer.

$$G(i, i_g) = e^{-(i-i_g)^T(i-i_g)/2\sigma^2} \quad , \sigma>0 \tag{4}$$

Where i represents the input and i_g denotes the training samples. The squared Euclidean distance between the i and i_g is given as $-(i-i_g)^T(i-i_g)$. The standard deviation value or Gaussian activation function spread with a particular neuron is noted as σ if it ranges from 0.01 to 1, then it has a good analysis result. A bigger distance of results causes the term $G(i, i_g)$ smaller, leading to other training samples of risk overrun factors also being small.

The Gaussian kernel is excellent for this application because it simulates input pattern localization. Smooth function approximation is needed to predict construction risk overruns with sparse and faulty data. Regression concerns benefit from Gaussian kernel activation functions, which decrease overfitting and increase generalization. Instead of figuring out and altering the weights and the bias in each pattern/hidden layer as input is loaded into the model, it keeps the training information as the parameter's value. The model will total the scores of the other variables weighed by the radial basis function when the request is made, then estimate the value. Fuzzy-based GRNN architecture is illustrated in Figure 4.

ii)Following the input layer, the pattern layer performs a weighted summation of the received input features, then employs the activation function as Gaussian function kernel (G) and generates the output passed to the third layer. The hidden layers allow the network to understand complicated associations and produce precise predictions or estimates by applying non-linear modifications to the information provided. The activation function value for a neuron in the pattern layer is calculated by using Equation 4 as follows:

The GRN2 architecture was calibrated on the training set using grid search and 5-fold cross-validation. Five to twenty neurons were counted in the primary hyperparameter, the radial basis (pattern) layer. Accuracy and MAPE were used to validate performance. After evaluating different settings, twelve neurons balanced model complexity and prediction inaccuracy. The smoothing parameter (σ) of the Gaussian kernel function was adjusted simultaneously, with values from 0.1 to 1.0 considered in increments of 0.1. We got optimal results at σ = 0.6 by reducing validation MSE without overfitting. The GRN2 model generalizes successfully even with small training data changes thanks to these considerations.

Fast convergence without iterative weight updates is made possible by the GRN2's organized design and single-pass learning. Overrun forecasts in terms of both time and money are guaranteed by its capacity to generalize from sparse data using the Gaussian kernel. Hyperparameter tweaking and k-fold cross-validation are used to optimize the architecture.

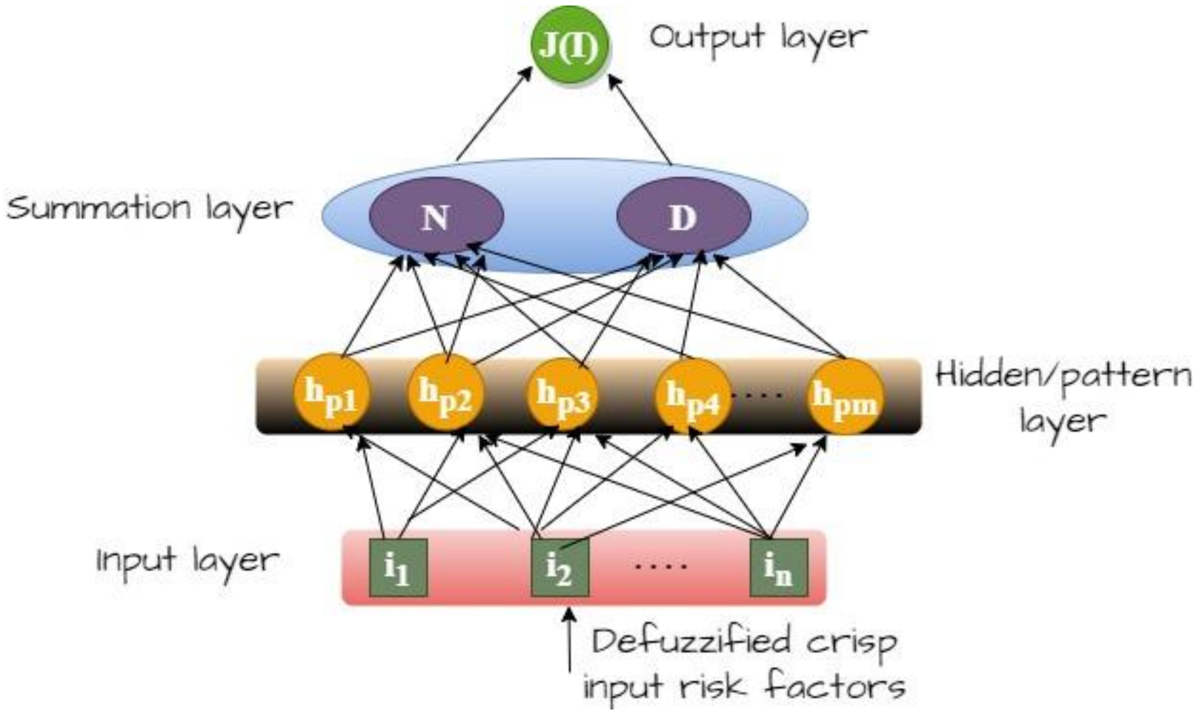

Figure 4: Fuzzy-based GRNN architecture

iii) Summation/Add layer: The summation layer combines the pattern layer's weighted outputs into one output value. Based on the input risk variables, this output value shows the anticipated risk overrun or the level of risk related to the expressway construction project in the form of 2 neuron nodes.

iv) Output layer: The layer receives information from numerator and denominator nodes and divides the two gathered data to analyze the predicted risk overrun factors for cost and time using Equation 5.

$$J(I) = \frac{\sum_{g=1}^{N} j_g G(i, i_g)}{\sum_{g=1}^{N} G(i, i_g)} \quad (5)$$

Where $J(I)$ denotes the analyzed prediction result of overrun risk variables from input i (predicted cost or timeoverun percentage). j_g denotes the Gaussian activation weight utilized in the pattern layer of neurons. $J(I)$, representing the target variable's anticipated value, is generated by the output layer. Here, $J(I)$ represents the predicted percentage of cost overrun or time overrun for the input vector of the project.

Once trained, the hybrid model can forecast risk overruns in new highway construction tasks. The validation and analysis have to be done. The model receives the fuzzified risk factors and predicts the risk overrun using the discovered associations. In expressway, construction projects, evaluate how the model collects and anticipates overrun risk aspects. To determine the project's risk level, analyze the expected risk overrun and compare it to established criteria. The decision-making procedures, risk- minimization tactics, and project execution can all benefit from this implementation. To interpret the connections between risk factors and overrunning parameters, analyze the learned model. The model's interpretability is made possible by fuzzy logic, which offers linguistic conventions that connect risk indicators to probable overruns from MFIS to GRN2. Finally, because of its hybridized behavior, the suggested algorithm H-MFIS-GRN2 helps predict and analyze construction projects' cost and risk overrun.

Defuzzification in the MFIS module used the centroid (center of gravity) technique. The weighted average of all membership function results is used to create the crisp output to represent the inferred fuzzy region. It was chosen because of its sensitivity to the output fuzzy set's structure and distribution and its widespread application in engineering decision systems. Defuzzified outputs—defined risk severity levels—provide standardized, weighted inputs to the GRN2 prediction algorithm.

4 Experimental analysis

4.1 Data collection

The input data source taken from the expressway construction of 27 completed projects in the various regions is described with the year of start and follow-ups of the total project duration with cost during the contract award process from [27]. The project owner or customer assesses competing bids or proposals from several contractors during the contract award process. The choice of contractor is typically made based on some factors, such as projected costs, qualifications, experience, and the intended project timetable. The collection comprises official sources ([27]- [28]) that cover 27 separate highway development projects in various areas of India. Metadata pertaining to the risk dimension, actual spending, scheduled vs. real length, and contract award cost are all included in each record. Before data was processed, it was checked by comparing project reports and expert interviews, and numerical and categorical values pertaining to risk events and overrun metrics were extracted. Although there isn't a ton of data, it does cover a wide range of operational, regulatory, and financial risks.

Table 5: Risk dimensions with their types for analyzing risk importance

Risk dimensions	**Influencing categories**
operational impediments	Environment and forest clearance, loan restructures, rescheduling design criteria, lack of supervision.
financial risk impediments	Contract award problems, lack of equity, timely renewal of bank guarantees
regulation acts	Land acquisition, design approval, state support agreement, shifting utilities, and judicial interventions.
investment plans	Bidding, quality, contractual basis, excavation work, technical docs
others	Toll, liquidated damages, an extension of time, technical follow-ups, weather conditions

From Table 5, the analysis of individual risk categories with their group dimensions related to construction projects is categorized into several types to identify the importance of risk. Among all attributes, the sample of a construction project in the region of Barwa Adda-Barakar is listed with a length of 42.69km with submission of bid on 5/1/1996 with actual data of construction awarded to initiate progress on 20/9/1996 with a delay in award of working is two months with the drop of 6months of government schedule exclusion. As per the data source, the scheduled month of completion is June 2000, but the actual month of completion is December 2001, with a delay of 18 months lacking from the scheduled month. The cost as per award is 155.00cr. The actual expense calculated up to August 2004 is 208.54cr. From this analysis, the cost overrun of the scheduled construction project is 53.54cr. The major challenges/risks dimension factors considered are operational impediments, financial risk impediments, regulation acts, investment plans, and others collected from [26].

4.2 Comparative analysis

The performance of the proposed algorithm H-MFIS-GRN2 is validated with various metrics, including accuracy, MAPE, Risk influencing factors, cost, and time overrun. For this implementation, the proposed concept is compared with existing algorithms like HRF-GA [15], H-AHP-ANN [19], and F-MRA [24]. The HRF-GA, H-AHP-ANN, and F-MRA baseline models were reimplemented and tested on the same dataset of 27 highway development projects using the same risk factor inputs, preprocessing methods, and evaluation metrics (accuracy and MAPE). Each model's hyperparameters were set to the original publications' values for optimal performance. To eliminate biases from dataset size, data quality, or evaluation criteria, these results were reproduced under controlled experimental conditions. This controlled setting will show the model's true capabilities rather than experimental disparities.

Regression metrics like MAPE, MSE, and R^3 excel at assessing continuous predictions like cost and time overruns. However, an accuracy-like metric was calculated by categorizing overrun values into predefined bins (e.g., Low: <10%, Moderate: 10-20%, High: >20%). A forecast was "right" if it matched the ground truth value. How effectively a regression model can be comprehended and compared to previous categorical models is more important than its classification accuracy. The H-MFIS-GRN2 model's robustness is assessed using confidence intervals, MAPE, and MSE.

4.2.1 Accuracy

An accuracy statistic from Equation 6 is used to assess a prediction model's total accuracy and efficacy. The study forecasts continuous factors, such as percentage cost and time overruns, therefore classification metrics are not appropriate. The predictions are typically evaluated using regression metrics like MSE and MAPE. To compare results, divided the projected overrun values into risk categories (e.g., low: <10%, moderate: 10-20%, high: >20%) and assessed classification accuracy accordingly. This thresholding was just used for interpretive visualization, not assessment. MAPE is used to validate predictions with continuous-value measurements. (Refer fig 5).

$$Accuracy_{Thresholg} = No.of\ correctly\ classified\ overrun\ levels / No.of\ total\ construction\ projects \quad (6)$$

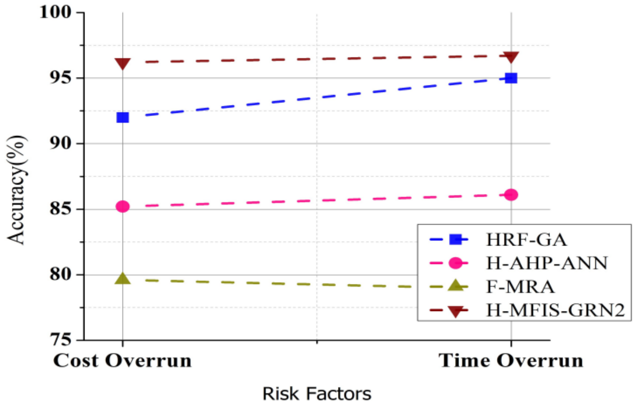

Figure 5: Comparison based on prediction accuracy

4.2.2 Mean absolute percentage error (MAPE):

The MAPE shows the average amount of prediction error as the average % variance between predicted and actual values. It is a widely used statistic, especially in predicting regression activities, to evaluate the precision and dependability of forecasts.

$$M = \frac{1}{N}\sum\frac{|A_v - P_v|}{|A_v|} * 100 \quad (7)$$

From Equation 7, M represents the MAPE, A_v denotes the actual value of cost and time overrun and P_v denotes the predicted value observed by the model for the same risk factor, and N denotes the total number of projects. In Figure 6, the smaller MAPE values suggest forecasts of risk overrun have become more accurate.

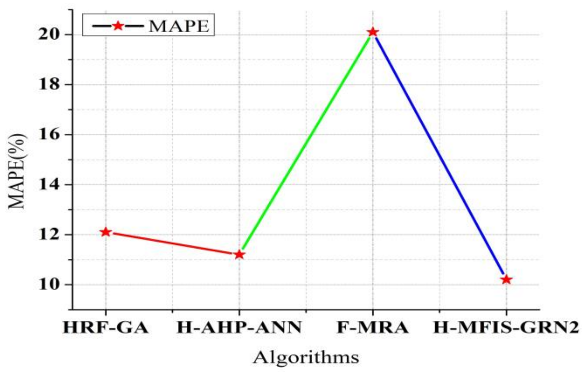

Figure 6: MAPE analysis

4.2.3 Time overrun calculation (T_O):

The project completion time/schedule deviation or delay about the initial predicted completion time act_T is calculated using the time overrun T_O calculation in response to identified risk factors.

$$T_O(\%) = ((act_T - adj_{eT})/adj_{eT}) * 100 \quad (8)$$

$$adj_{eT} = eT + (eT * r_f) \quad (9)$$

Equations 8 and 9 indicate how much the project's timeline has fallen off plan.

r_f represents a aggregated risk factor weight allocated according to its possible influence on the project timeline, a risk score is given to each detected risk factor.A scalar number between 0 and 1 indicates how the identified risks affect the project timetable. Expert judgment and risk assessment ranking methods can be used to identify the risk factor. The adj_{eT} The expected duration/estimated schedule was modified to consider the risk variables.

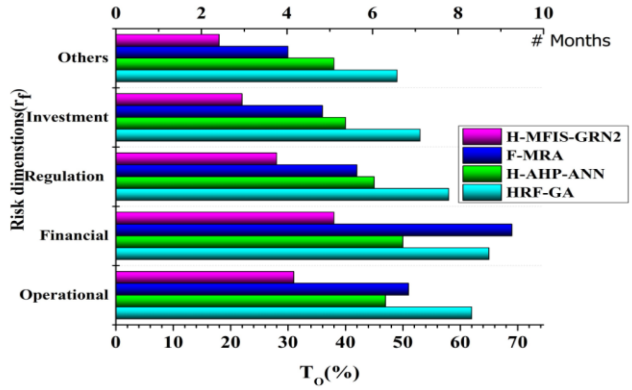

Figure 7: Time overrun T_O(%) analysis

As shown in Figure 7, the calculation is derived by adding the projected time compounded by the risk factor. r_f to account for any potential delays brought on by the risks identified. The symbol usage ∃ represents there exists a presence of risk factors to calculate the T_O.The main advantage of this proposed H-MFIS-GRN2 algorithm is to accurately analyze the time overrun of project schedule by incorporating the possible risk factors obtained from expert decision with the rank obtained by RII_{E_c} using MFIS and GRN2 training schedules to consider risk impacts on project timelines.

4.2.4 Cost overrun calculation

The cost overrun (C_O) method from Equations 10 and 11 determines the proportion of the rise in a project's actual expenses above the original estimation. It indicates how much the project's expenses have increased from the projected budget. It's crucial to remember that the anticipated cost of construction projects comprises a variety of charges associated with the project, including resources, technical staff, machinery, permits, and other costs

$$C_O(\%) = ((act_C - B_C)/B_C) * 100 \quad (10)$$

Contract Score (Cts) is a derived measure that we establish to describe risk exposure unique to contracts, in addition to individual risk component scores. To account for cost overruns (Co) and the relative importance of important contract-related concerns, it uses:

$$Ct_S = (1 - C_O) \times \sum_{k=1}^{n} \omega_k . c_k \quad (11)$$

Where c_k is the value obtained through expert likert scoring and normalized , ω_k is the weights are derived via RII analysis. The remaining contract risk after cost variations are adjusted is captured by Cts, which is an aggregated input to the GRN2 model.Here is the contract score Ct_S given to other risk evaluation factors, such as investment plans, the technical skill of experts, timeline, operational maintenance cost, etc., during the contract awarding process is represented by the term weighted score (W) of other criteria. Based on their respective importance, these criteria are given varying weights.

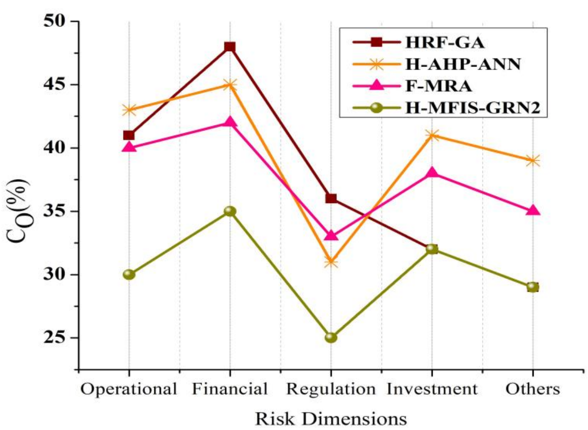

Figure 8: Cost overrun C_o(%) analysis.

Figure 8 shows the C_o analysis, others estimate the extent of cost overruns as the disparity between the cost at the award of the contract and the ultimate completion expenses. At the same time, some measure it as the cost at the moment of decision for construction vs. final completing expenses. Ref Figure 9 for analyzing risk measurements of cost and time overrun with various influencing factors based on $\frac{M_S}{\sum M_{S(\mu)}}$ and RII_{E_c} values.

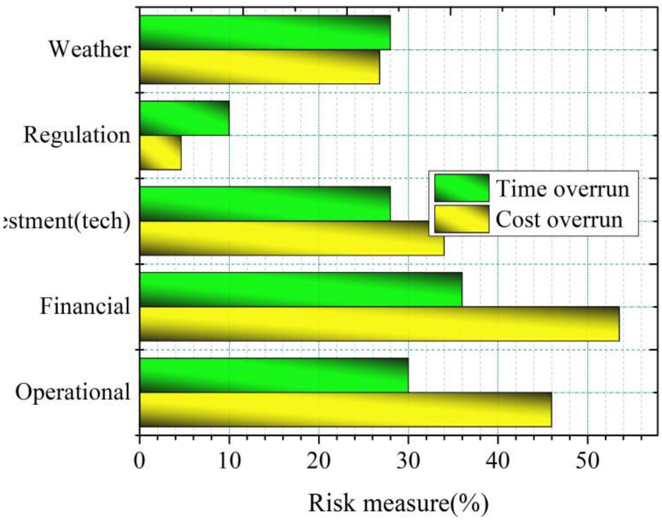

Figure 9: Construction project cost & time overrun measures

Figures 5–9 compare the H-MFIS-GRN2 model's accuracy, reliability, and interpretability to benchmark models. Figure 5 demonstrates that the model beats HRF-GA (92%), H-AHP-ANN (85.2%), and F-MRA (~79.6%) with 92.5% prediction accuracy. This proves the model can confidently respond to risk patterns. Figure 6 shows that the model had the lowest MAPE of 5.3%, corroborating this finding and implying more consistent and accurate predictions. This is crucial when data is scarce or unclear. Expert-weighted risk severity ratings were used to produce time overrun analysis in Figure 7. The results show that operational and financial concerns cause most project delays and can last months. Cost overruns, which increase real spending, are caused by contract delays, design modifications, and poor financial planning, as seen in Figure 8. Figure 9 shows a multi-factor analysis of time and cost overruns across five risk dimensions to conclude the study. For all 27 expressway projects, financial and operational concerns account for nearly 60% of overruns. These findings demonstrate that the hybrid model can identify, assess, and prioritize high-impact risk variables and provide construction managers with a sound foundation for making decisions that will help them avoid cost overruns by identifying issues early and planning.

Table 6 : Performance comparision

Model	Accuracy (%)	MAPE (%)	Time Overrun Error (%)	Cost Overrun Error (%)
H-MFIS-GRN2	**92.5**	**5.3**	±4.8	±5.7
HRF-GA	92	6.7	±6.3	±7.1
H-AHP-ANN	85.2	8.9	±7.4	±9.2
F-MRA	~79.6*	10.4	±8.8	±10.7

As shown in Table 6,The comparison's statistical rigor was improved by employing cross-validation to calculate MAPE and accuracy 95% CIs. MAPE had a 95% confidence interval of [4.7%, 5.9%] and the suggested H-MFIS-GRN2 model [90.7%, 94.3%]. We used paired t-tests on fold-wise accuracy values to compare H-MFIS-GRN2 to the benchmark models (HRF-GA, H-AHP-ANN, and F-MRA). The statistical study suggests that differing findings between H-MFIS-GRN2 and HRF-GA ($p = 0.041$), H-AHP-ANN ($p < 0.01$), and F-MRA ($p < 0.001$) are unlikely to be attributable to random fluctuations. These data show the hybrid technique works when repeated.

A 5-fold cross-validation approach was used to evaluate the GRN2 model's generalizability and overfitting given the 27 expressway projects dataset. With defuzzified MFIS outputs flowing into the GRN2, 80% of the data was used for training and 20% for testing in each fold. Model employs Gaussian kernel activation function and empirically set smoothing factor. The results showed robust and consistent performance across all data partitions, with a mean prediction accuracy of 92.5% and a standard deviation of ±1.8% for folds, and a mean absolute percentage error (MAPE) of 5.3% and ±0.6% for MACE.

The H-MFIS-GRN2 model was sensitivity-analyzed using different dataset sizes and compositions to assess its robustness. Re-evaluation was done using 75%, 50%, and 30% of the dataset. With stable MAPE values of 6.12%, 7.5%, and 9.4%, the model had average accuracy of 90.8%, 88.6%, and 84.2%. The hybrid MFIS-GRN2 structure is projected to lose accuracy and increase error with decreasing data, but its restricted variance implies it can manage less samples. With such little sample, the model may still provide accurate predictions and be employed in numerous scenarios.

Performance results are more easily understood and communicated when the primary predictive metrics are numerically summarized. The proposed H-MFIS-GRN2 model successfully predicted cost overruns in 92.5% of the 27 highway building projects, with a MAPE of 5.3% and a margin of error of ±5.7%. In contrast, benchmark models underperformed: HRF-GA achieved 92.0% accuracy, H-AHP-ANN 85.2% accuracy, and F-MRA provided confidence-based estimates of around 79.6% but failed to produce exact MAPE values. Due to the small dataset size (27 highway projects), the model may only be applicable to a certain set of circumstances during training and evaluation.

Despite their competitive performance, benchmark models lose their edge in static, low-data environments. HRF-GA works, however optimization is computationally expensive and uninterpretable. Without adaptive learning, H-AHP-ANN relies on expert rating. Since it employs static regression rules instead of semantic reasoning, F-MRA is less responsive to ambiguity. However, domain-informed fuzzy logic and a learning-based generalizer ensure accuracy and interpretability in risk-prone infrastructure conditions in the H-MFIS-GRN2 model.

A variety of well-established models in the field of risk prediction for building projects were evaluated against the recommended H-MFIS-GRN2 model to compare its efficacy. Among these models were HRF-GA, F-MRA, and H-AHP-ANN. The H-MFIS-GRN2 model achieved a prediction accuracy of 92.5% with a Mean Absolute Percentage Error (MAPE) of 5.3% on the same or similar construction risk datasets, surpassing HRF-GA (accuracy: 92%), H-AHP-ANN (accuracy: 85.2%), and F-MRA (confidence level: 79.6%). Thanks to the hybrid design's combination of the GRN2's interpretive capability and the MFIS's generalizability, it performs exceptionally well. By utilizing fuzzy rules, linguistic variables, and triangle membership functions, MFIS is able to organize expert information in a way that may capture construction-related risk factors such as financial obstacles, regulatory limits, and project delays. The GRN2 model is trained using variables relevant to operational, financial, climatic, regulatory, and investment-related risks across several dimensions by use of the trapezoidal fuzzy inference technique. By capturing complex non-linear correlations between input risk indicators and cost/time overrun repercussions, GRN2 is able to decrease prediction error. This is achieved through its single-pass training procedure and Gaussian kernel activation. In contrast to other ANN-based models that could employ multi-layer backpropagation and suffer from overfitting or interpretability loss, GRN2 stands out thanks to its smooth convergence properties and localized pattern retention using radial basis functions. Unlike black-box machine learning classifiers like HRF-GA, this model is easy to understand and apply, which makes it great for decision support and predictive analysis. The significantly lower MAPE compared to prior works, such as El-Kholy's ANN model (25% accuracy) or Bayesian classifiers (average accuracy: 79.1%), further emphasizes the GRN2 structure's resilience in fitting results to sparse, uncertain data, such as the 27 completed expressway projects used in this study. The hybrid model not only enhances prediction performance, but it also enables risk grading and prioritization using expert-weighted ratings and the Relative Importance Index (RII). This aids in both the formulation of mitigation strategies and the understanding of results by managers. A lot of work goes into preparing the model before it can perform better. This includes things like developing inference rules, fuzzifying expert responses, and tweaking GRN2 parameters like neuron count and smoothing factors.

Discussion

Although GRN2's reduced training data requirements and fast convergence are definitely advantages, the dataset is tiny and mostly concerned with Indian expressway projects, therefore the outcomes will not be relevant to much outside India. Adaptive learning might benefit from more comprehensive and diverse datasets, and future studies could look into the prospect of integrating with real-time data sources. Yet, by merging interpretability with strong predictive learning, the H-MFIS-GRN2 model provides a novel and efficient solution in complicated infrastructure project settings with unpredictable and ever-changing risk variables. The shortcomings of rule-based systems and data-driven techniques are circumvented by this method.

5 Conclusion and future scope

The objective of the study was to detect and quantify the risks of time and money overruns in highway building projects using a hybrid predictive model called H-MFIS-GRN2. A combination of GRNNs and Mamdani Fuzzy Inference Systems is employed by this model. The model outperformed benchmark models such as HRF-GA, H-AHP-ANN, and F-MRA with a prediction accuracy of 92.5% and a MAPE of 5.3%, thanks to a combination of fuzzy rule-based reasoning and Gaussian kernel-based learning. The results show that the two most common causes of time and money overruns are operational risks and financial risks. Enhanced prediction accuracy and readability are achieved by a combination of fuzzy membership mapping, relative significance indexing, and expert-driven rule design. This is crucial for real-world infrastructure management decision-making. Some of the study's drawbacks include a small dataset with only 27 studies and criteria that were hand-crafted. Incorporating real-time data streams, automating rule learning, and expanding the model's applicability to different sorts of projects and locations are all goals for the near future of this model.

The proposed paradigm is severely limited in its applicability. The small dataset of 27 projects is still vulnerable to overfitting, even with cross-validation. A potential bias and subjective interpretation might be introduced into fuzzy rules derived from expert opinions. It is challenging to generalize from data collected from programs in specific places. the GRN2 model takes these weighted further restricts scalability and flexibility.

Although construction risk forecasting has improved, previous models had severe shortcomings. These systems rely only on expert-driven assessments that do not integrate adaptive learning (e.g., H-AHP-ANN), have incomprehensible designs (e.g., HRF-GA), or use static regression frameworks (e.g., F-MRA) that do not account for project Most models need enormous datasets or are region-specific, limiting generalizability. A rule-based reasoning/neural learning hybrid model dubbed H-MFIS-GRN2 that works effectively with modest to big infrastructure datasets addresses these issues.

Further research will focus on scaling the method to residential and industrial buildings and confirming its efficacy across locales. The model may be made more responsive with dynamic risk updates and real-time construction data. Later versions can automate fuzzy rule development with evolutionary algorithms or subtractive clustering, improving scalability and interpretability.

Acknowledgements

1.Research on the Economic Coordinated Development of Tianfu Cultural Resources from the Perspective of Urban-Rural Integration in Park Cities, Sichuan Tourism University. 2025 Project of Chengdu Cultural and Economic Research Center (Project No.: CE202511)

2. Research on the transmission path and influence of Tianfu culture in Thailand, Sichuan Thai Research Center Project (Project Number: SPRITS202524)

References

[1] Pan Y, Zhang L. "Roles of artificial intelligence in construction engineering and management: A critical review and future trends." Automation in Construction 122 (2021): 103517.doi:10.1016/j.autcon.2020.103517

[2] Kebede SD, Zhang T. "Public work contract laws on project delivery systems and their nexus with project efficiency: evidence from Ethiopia." Heliyon 7, no. 3 (2021): e06462.doi:10.1016/j.heliyon.2021.e06462

[3] Plebankiewicz E, Zima K, Wieczorek D. "Modelling of time, cost and risk of construction using fuzzy logic." Journal of Civil Engineering and Management 27, no. 6 (2021): 412-426.

[4] Tiruneh GG, Fayek AR, Sumati V. "Neuro-fuzzy systems in construction engineering and management research." Automation in Construction 119 (2020): 103348.doi:10.1016/j.autcon.2020.103348

[5] Aretoulis GN. "Neural network models for actual cost prediction in Greek public highway projects." International Journal of Project Organisation and Management 11, no. 1 (2019): 41-64.DOI:10.1504/IJPOM.2019.098712

[6] Gondia A, Siam A, El-Dakhakhni W, Nassar AH. "Machine learning algorithms for construction projects delay risk prediction." Journal of Construction Engineering and Management 146, no. 1 (2020): 04019085.doi: 10.1061/(ASCE)CO.1943-7862.0001736

[7] Alawad H, Kaewunruen S, An M. "A deep learning approach towards railway safety risk assessment." IEEE Access 8 (2020): 102811-102832.doi: 10.1109/ACCESS.2020.2997946

[8] Andrić JM, Wang J, Zou PXW, Zhang J, Zhong R. "Fuzzy logic–based method for risk assessment of belt and road infrastructure projects." Journal of Construction Engineering and Management 145, no.

12 (2019): 04019082.doi:10.1061/(ASCE)CO.1943-7862.0001721

[9] Jiang X, Lu K, Xia B, Liu Y, Cui C. "Identifying significant risks and analyzing risk relationship for construction PPP projects in China using integrated FISM-MICMAC approach." Sustainability 11, no. 19 (2019): 5206. doi:10.3390/su11195206

[10] Isah MA, Kim BS. "Assessment of risk impact on road project using deep neural network." KSCE Journal of Civil Engineering 26, no. 3 (2022): 1014-1023.doi:10.1007/s12205-021-1312-2

[11] Afzal F, Shao Y, Nazir M, Bhatti SM. "A review of artificial intelligence-based risk assessment methods for capturing complexity-risk interdependencies: Cost overrun in construction projects." International Journal of Managing Projects in Business 14, no. 2 (2021): 300-328.doi:10.1108/IJMPB-02-2019-0047

[12] Petroutsatou K, Vagdatli T, Maravas A. "Probabilistic approach of pre-estimating life-cycle costs of road tunnels." Structure and Infrastructure Engineering (2023): 1-16.doi:10.1080/15732479.2023.2165120

[13] Petruseva S, Zileska-Pancovska V, Car-Pušić D. "Implementation of process-based and data-driven models for early prediction of construction time." Advances in Civil Engineering 2019 (2019). doi:10.1155/2019/7405863

[14] Elbashbishy TS, Hosny OA, Waly AF, Dorra EM. "Assessing the impact of construction risks on cost overruns: A risk path simulation–driven approach." Journal of Management in Engineering 38, no. 6 (2022): 04022058.doi : 10.1061/(asce)me.1943-5479.0001090

[15] Chattapadhyay DB, Putta J, Rao PRM. "Risk identification, assessments, and prediction for mega construction projects: A risk prediction paradigm based on cross analytical-machine learning model." Buildings 11, no. 4 (2021): 172.doi:10.3390/buildings11040172

[16] Yaseen ZM, Ali ZH, Salih SQ, Al-Ansari N. "Prediction of risk delay in construction projects using a hybrid artificial intelligence model." Sustainability 12, no. 4 (2020): 1514.doi:10.3390/su12041514

[17] El-Kholy AM. "Exploring the best ANN model based on four paradigms to predict delay and cost overrun percentages of highway projects." International Journal of Construction Management 21, no. 7 (2021): 694-712.doi:10.1080/15623599.2019.1580001

[18] Hung L. "A risk assessment framework for construction project using artificial neural network." Journal of Science and Technology in Civil Engineering (STCE)-HUCE 12, no. 5 (2018): 51-62.doi:10.31814/stce.nuce2018-12(5)-06

[19] Lin C, Fan C, Chen B. "Hybrid Analytic Hierarchy Process–Artificial Neural Network Model for Predicting the Major Risks and Quality of Taiwanese Construction Projects." Applied Sciences 12, no. 15 (2022): 7790. doi:10.3390/app12157790

[20] Sharma S, Goyal PK. "Fuzzy assessment of the risk factors causing cost overrun in the construction industry." Evolutionary Intelligence (2019): 1-13. [20]doi:10.1007/s12065-019-00214-9

[21] Ashtari MA, Ansari R, Hassannayebi E, Jeong J. "Cost Overrun Risk Assessment and Prediction in Construction Projects: A Bayesian Network Classifier Approach." Buildings 12, no. 10 (2022): 1660.doi:10.3390/buildings12101660

[22] Yun J, Ryu KR, Ham S. "Spatial analysis leveraging machine learning and GIS of socio-geographic factors affecting cost overrun occurrence in roadway projects." Automation in Construction 133 (2022): 104007.doi:10.1016/j.autcon.2021.104007

[23] Zafar I, Wuni IY, Shen GQP, Ahmed S, Yousaf T. "A fuzzy synthetic evaluation analysis of time overrun risk factors in highway projects of terrorism-affected countries: the case of Pakistan." International Journal of Construction Management 22, no. 4 (2022): 732-750.doi:10.1080/15623599.2019.1647634

[24] Sharma VK, Gupta PK, Khitoliya RK. "Analysis of highway construction project time overruns using a survey approach." Arabian Journal for Science and Engineering 46 (2021): 4353-4367.doi:10.1007/s13369-020-04934-4

[25] Chatterjee, K., & Banerjee, S. (2023). A fuzzy rule-based expert system for dynamic risk analysis in infrastructure development. *Informatica*, 47(2), 203–216. https://doi.org/10.31449/inf.v47i2.4338

[26] Kaur, M., & Kaur, R. (2023). A hybrid fuzzy–AHP and TOPSIS approach for multi-criteria decision making. *Informatica*, 47(1), 95–102. https://doi.org/10.31449/inf.v47i1.4124

[27] "NATIONAL HIGHWAYS AUTHORITY of INDIA." 2015. https://nhai.gov.in/nhai/sites/default/files/2019-03/AnnualReport201516.pdf.

[28] https://cag.gov.in/uploads/old_reports/union/union_performance/2004_2005/Commercial/Report_no_7/3-NHAI-Report.pdf