International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

Adaptive Evolutionary Fuzzy Inference System for Accurate Software Project Planning and Cost Prediction

Dr. Ajay Jaiswal[1], Siddharth Singh[2], Trisha Soni[3], Shubh Khatri[4], Insha Khan[5]
[1] Professor, Department of Computer Science and Engineering,
Prestige Institute of Engineering Management and Research Indore
[2], [3], [4], [5] Student B.Tech, Department of Computer Science and Engineering,
Prestige Institute of Engineering Management and Research Indore

ABSTRACT
This research suggests a hybrid computational model of most appropriate software project planning cost estimation through the combination of fuzzy logic and genetic algorithm optimisation with machine learning models. The study uses three systematic datasets used to show different software development settings to test the predictive ability of the suggested method. Gene elimination and genetic optimisation, recursive feature elimination, data preprocessing and feature extraction are used to improve the relevance of features and dimensionality reduction. The fuzzy inference mechanisms are used to model the uncertainty of the project characteristics like complexity, duration and team experience and the genetic algorithm is used to optimise the membership functions and rule weights so that the error in estimation is reduced to a minimum. Experimental evidence confirms that the hybrid model is always better in lowering the values of MAE and RMSE and increasing the values of R 2. The results are that uncertainty-conscious reasoning with evolutionary optimisation leads to better prediction accuracy and robustness, which makes the model applicable in the cost estimation and decision support of software projects that are expected to be reliable even in the dynamic development settings.

Keywords- Software cost estimation, fuzzy logic, genetic algorithm, hybrid model, project planning, machine learning

1. INTRODUCTION
Proper cost estimation of software project planning is one of the most difficult and complicated issues of software engineering and project management. The uncertainty which surrounds the cost prediction is a very non-linear and uncertain issue because of the dynamic nature of software development which is characterised by changing requirements, diverse team experience, uncertainty in technology and the risk level which is affected by factors. Some of these traditional algorithmic methods of estimating costs like Constructive Cost Model (COCOMO) and regression-based estimation methods usually presuppose that there will be a linear relationship between project variables and thus will not work well in these software projects that can have inherent vagueness and ambiguity. This results in budget overruns, schedule slippage and inefficient allocation of resources as a result of inaccurate estimations, which eventually have an impact on project success and organisational competitiveness. In the past ten years, there has been a greater interest in intelligent computational methods that are able to deal more efficiently with uncertainty, nonlinear interactions of project attributes. Here,

fuzzy logic has become a viable paradigm because it is capable of approximating imprecise and natural information to allow expert judgement and subjective evaluation to be incorporated into cost estimation models (Jorgensen, 2014). At the same time, evolutionary optimisation algorithms (genetic algorithms) have shown strong potential in optimisation of the parameters of the model and exploration of the best solutions in multidimensional and complex problem space; hence providing a powerful means of optimising the accuracy of the estimation (Harman, 2012).

Fuzzy logic is the incorporation of the genetic algorithm in fuzzy logic offers a synergistic framework that has the potential to overcome the weakness of traditional estimation approaches. The use of fuzzy logic of uncertain project features like project complexity, project requirement stability and team ability is represented using using linguistic variables and rule based inferences systems to enhance the interpretability and flexibility of estimation models. Nonetheless, the success of the fuzzy systems is highly contingent on the proper design of the membership functions and inference rules which are not always easy to tune by hand. These parameters can be optimised automatically using genetic algorithms to mimic the process of evolution in nature, such as selection, crossover and mutation to find parameter combinations that minimise the error of estimation. This hybridisation can be used to create adaptive and self-tuning estimation models that can learn based on past project history and at the same time, it can incorporate human knowledge. Empirical results of software cost estimation have revealed that hybrid intelligent models are better than other standalone statistical and machine learning methods as they generate lower prediction error and greater generalisation ability in a varying project environment (Malik et al., 2015; Shepperd and MacDonell, 2012). This suggests that evolutionary fuzzy systems have a lot of potential in enhancing the decision-making process within the software project planning context especially in situations that are characterised by uncertainty and lack of information.

Besides dealing with uncertainty, other interacting variables should be considered as the optimal project planning that must include project size, duration of the project development, experience of the team, novelty of the technology and organisational limitations. These variables tend to have complicated nonlinear relations which are hard to model under strictly deterministic techniques. The genetic algorithms offer a global optimisation scheme that has the ability of searching high search space and bypassing local minimum which is particularly valuable in the estimation of high-dimensional problems. They in combination with fuzzy inference mechanisms allow building strong predictive models that are both interpretable and efficient in computation. The recent developments in hybrid soft computing methods have underscored their use in other areas of the software engineering process such as effort prediction, risk assessment and resource allocation which further emphasize their suitability in contemporary project management procedures (Aleti and Moser, 2013; Mittas and Angelis, 2013). In spite of these achievements, it is still necessary to conduct a thorough research that will bring together fuzzy reasoning and genetic optimisation in a coherent framework that will be able to estimate the cost of software projects and be validated empirically on organised datasets representing real-world development processes. The current research paper is informed by this research gap and seeks to provide an intelligent hybrid framework, which would utilise the strengths of the fuzzy logic and genetic algorithm in order to improve the accuracy, adaptability and robustness of software project cost estimation.

2. IMPORTANCE OF THE STUDY

The significance of this work is that it will make software project planning estimation of the costs more reliable and precise by integrating the use of fuzzy logic and genetic algorithms. Effective cost estimation is a pillar of a successful software project management as it directly reflects on budget, schedule, allocation of resources and confidence among the stakeholders. Under modern software development conditions, where there is a fast-paced technological change, *distributed teams, and constantly shifting requirements, traditional estimation methods do not usually resolve uncertainty and inaccuracy in the project characteristics. This paper will counter that weakness by suggesting a hybrid computational model that can model both the vague and linguistic variables and at the same time optimise the parameters of the estimations through evolution search mechanisms. This is significant especially considering that this approach improves the decision making ability of the project managers since they can make more realistic and flexible predictions of the costs reducing the risk of cost overruns and project failure.

In theoretical terms, the paper enhances the area of knowledge in estimation of software engineering by illustrating the synergistic fuzzy reasoning and genetic optimisation value. With the assistance of fuzzy logic, subjective human judgement and vague project description variables like complexity, risk and team capability can be represented, which are hard to measure strictly with the help of the traditional mathematical framework. The performance of fuzzy systems however, is greatly dependent on the appropriate arrangement of the membership functions and inference rules. This study adds into consideration genetic algorithms and presents an automated optimisation mechanism which is used to evolve these parameters to obtain a better predictive performance. This integration also helps in the formulation of adaptive estimation models which are capable of learning using past project information whilst being interpretable, a feature that is highly unattainable in pure machine learning that is driven by data. The research, therefore, provides a methodological development slotting between expert and data-driven estimation paradigms, which has been identified as an essential requirement of modern software cost estimation studies (Jorgensen, 2014; Harman, 2012).

The practical significance of the study is also immense especially to the software development organisations that work in competitive and resource limited environment. Cost estimation may be inaccurate resulting in huge losses of funds, poor use of human resources and lack of trust by the clients. The study will assist organisations to make more effective project planning and monitoring by enabling them to predict possible cost deviations, and preempt the adjustment of strategies. Moreover, the evolutionary optimisation approach makes the suggested model scalable and adaptive to a variety of project conditions, such as agile and hybrid development strategies. Such flexibility is essential in current software engineering, where the characteristics of a project often change throughout the lifecycle of project development. As such, the research is practical to both academics and industry professionals who need to have well-founded, adaptable and data-driven techniques to estimate the cost of software projects, which will eventually lead to the success of the project and the resulting better organisational performance.

3. THEORETICAL AND CONTEXTUAL CONTRIBUTION OF THE RESEARCH

The theoretical value of this study is based on the idea of developing hybrid soft computing approaches to software project cost estimation, especially based on the orderly combination of fuzzy and genetic algorithms and modelling approach. Theoretical models that are currently in use in software cost estimation have been characterized on a large scale by parametric methods and statistical regression methods, which presuppose that there exists a deterministic relationship between the project variables and do not always well account for the ambiguity and subjectivity of the real world development environment. This paper builds upon the theoretical foundation of the relationship between linguist and imprecise project characteristics, including requirement volatility, technical complexity and staff competency, through the formal representation and processing through a mathematical inference system by integrating fuzzy logic into the estimates. This adds to the theoretical discussion of the uncertainty modelling in software engineering as it illustrates how fuzzy sets and rule-based reasoning can be used to hedge the gap between the qualitative managerial judgement model and the quantitative prediction model. Moreover, the additive use of genetic algorithms allows the inclusion of an evolutionary optimisation aspect to the theoretical basis of all the estimation models to allow dynamic tuning of membership functions and rule weights in order to obtain better predictive power and generalisation with different datasets. This hybridisation is consistent with the current theoretical trends in computational intelligence, which promote the integration of interpretability and flexibility in solving complex problems of software project management (Harman, 2012; Jorgensen, 2014).

Besides the theoretical expansion, the study also provides a contextual contribution to theoretical research by putting the hybrid estimation framework into the context of the reality of contemporary software development ecosystems. The environments in which contemporary projects take place are characterised by the fast rate of requirement changes, distributed cooperation, the presence of technological heterogeneity and the increased market pressures that result in the significant uncertainty in the planning and budgeting functions. The contextual relevance of this study is the fact that these dynamic relevant conditions can be modelled by using flexible fuzzy inference mechanisms but variable estimation parameters can be optimised using genetic algorithms as project data changes. Such a contextual fit makes sure that the proposed model is not still a mere theoretical abstraction, but rather a practically implementable tool of decision-support which has the potential to mitigate the estimation problem of software organisations working either in traditional or agile development environments. Further, the methodological framework of the study puts it into context with structured datasets that represent the empirical attributes of real-world projects including size, complexity, duration and team experience and thus empowers the empirical basis of the study and increases its probability of being applicable in different organisational settings.

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

The paper is also valuable to the overall picture of intelligent project planning, showing how the hybrid evolutionary fuzzy systems may be used as a mediator between the expert knowledge and the information-based analytics. Cost estimation in most organisation set ups remains largely subject to expert guess work which despite its merits is not consistent and is not easy to normalise. The research offers a means of knowledge capture of tacit managerial knowledge by formalising expert reasoning using fuzzy rules, and optimisation of the fuzzy rules using genetic algorithms in a reproducible and scalable way. This contribution within the context is most important in the case of organisations that want to institutionalise best practices in project planning and cost control and use past project repositories to improve on that. By extension, the study does not only enhance the theoretical environment of software cost estimation, but it also provides a context sensitive framework, responsive to the increasingly dynamic complexities of the modern software development cycle, hence making it possible to enhance the viability of more adaptive and evidence informed project planning practices.

4. LITERATURE REVIEW

On one hand, the software project cost estimation domain has undergone radical transformation during the last two decades and leans towards the non-parametric and non-regressive methods and, on the other hand, intelligent and hybrid methods. The earliest models such as the one of the COCOMO and other forms of statistical estimation were highly reliant on historical data and deterministic mathematical association among the size of the software and the quantity of effort required to create software. Although these models provided a more systematised method of estimation, researchers realised that they were not good when uncertainty would be incorporated, volatile requirement and nonlinear relationships between the project characteristics. Jogensen and Shepperd (2011) note that the correctness of estimation is incredibly situational and there is no model that is superior to the rest in different project settings. This realization evoked the search of the soft computing techniques that could resemble ambiguity and inaccuracy of the software engineering data. Consequently, fuzzy logic was suggested as a high-profile research focus because of the ability to model linguistic and uncertain data, and, therefore, subjective management decision and imprecise project definitions to be incorporated in the formal model of computations.

Fuzzy logic estimation model has been widely researched because it has the potential of handling inaccurate inputs of the estimation model such as the complexity of the project, experience and stability of the requirements among the team members. Fuzzy based techniques of analogy estimation can effectively function with both heterogeneous and homogeneous data (numerical and categorical) as demonstrated by Azzeh, Neagu and Cowling (2010) and in so doing, can achieve successful prediction outcomes than the other traditional techniques of analogies. Still on the same note, Nassif, Capretz and Ho (2012) came up with a fuzzy logic model to perform software effort estimation and found that linguistic variables and fuzzy inference rules enhanced the realism in estimations and reduced prediction error. Further empirical results of Mendes and Mosley (2013) found out that the fuzzy models when applied most, are cases when an early information of a project is not comprehensive or probabilistic since the models can be applied in the revisions of estimates as more information is known. All these investigations confirm the fact that fuzzy reasoning provides an adaptable method to the modelling of uncertainty and subjectivity in the software project planning to improve interpretability and more practical usefulness of the estimation results.

Despite these advantages, the effectiveness of fuzzy inference systems is very much dependent on a well-valued designing of the membership functions and rule bases which in most instances is determined by the expertise. They may be manually optimized to add bias to their parameters and make models less scalable across a range of project contexts. In order to solve this deficiency, researchers have increasingly resorted to the application of evolutionary optimisation algorithms, and in particular to genetic algorithms, to automatically optimise fuzzy systems. Harman (2012) emphasized on search based software engineering algorithms like genetic algorithm, which is efficient in solving multidimensional and complex problems in software engineering. The use of genetic algorithms to fuzzy membership functions in cost estimation situations enables optimisation of membership functions, rule weight and selection of the most appropriate input features, which improves the quality of the predictions and generalisation. Studies such as Huang, Chiu and Lee (2013) found that genetic-fuzzy models are much more effective than single fuzzy systems, and conventional regression models in evaluating real world project data since they share the capacity to compute the optimal parameter settings through an evolutionary search.

An integration between genetic algorithms and the fuzzy logic has therefore become one of the dominant hybrid techniques in software cost estimation studies. Malik, Singh and Kaur (2015) studied a hybrid fuzzy-genetic model and experienced a great reduction in the mean values of the relative error compared to the classical estimation techniques. According to their findings, evolutionary tuning will enhance the adaptability of fuzzy systems to varied project set ups with nonlinear and uncertain relationships between cost drivers. This opinion was also reflected by Mittas and Angelis (2013) who emphasized on the fact that evolutionary computer techniques provide effective ways of exploring large regions and finding the most suitable parameter estimates of the predictive models. These types of optimisation functions are particularly applicable in the estimation of software engineering where the interactions between the characteristics of a project is complex and in many cases poorly understood.

In addition to the literature on the area of fuzzygenetic, a broader variety of hybrid intelligent systems have been examined by other authors, which merge fuzzy reasoning with machine learning methods such as neural networks and support vector machines. Aleti and Moser (2013) observed that the hybrid intelligent models enhance the precision of the estimates since they take into account the patterns of numbers as well as the qualitative expertise knowledge to enhance the strength of the model in the different project scenarios. Similarly, Kocaguneli, Menzies and Keung (2012) have observed that there exists a need to fight the variability and uncertainty which follow the process of software effort estimation, alongside the data-driven learning and domain knowledge. It has been demonstrated that the performance of the hybrid models is superior to that of their respective parts; machine learning techniques have been proved to be good in pattern recognition but fuzzy logic techniques have been proven to be interpretable and capable of dealing with uncertainty. The literature, in its turn, is inclined to suggest the development of holistic frameworks that would bring about the balance in the data-driven optimisation and the anthropocentric reason.

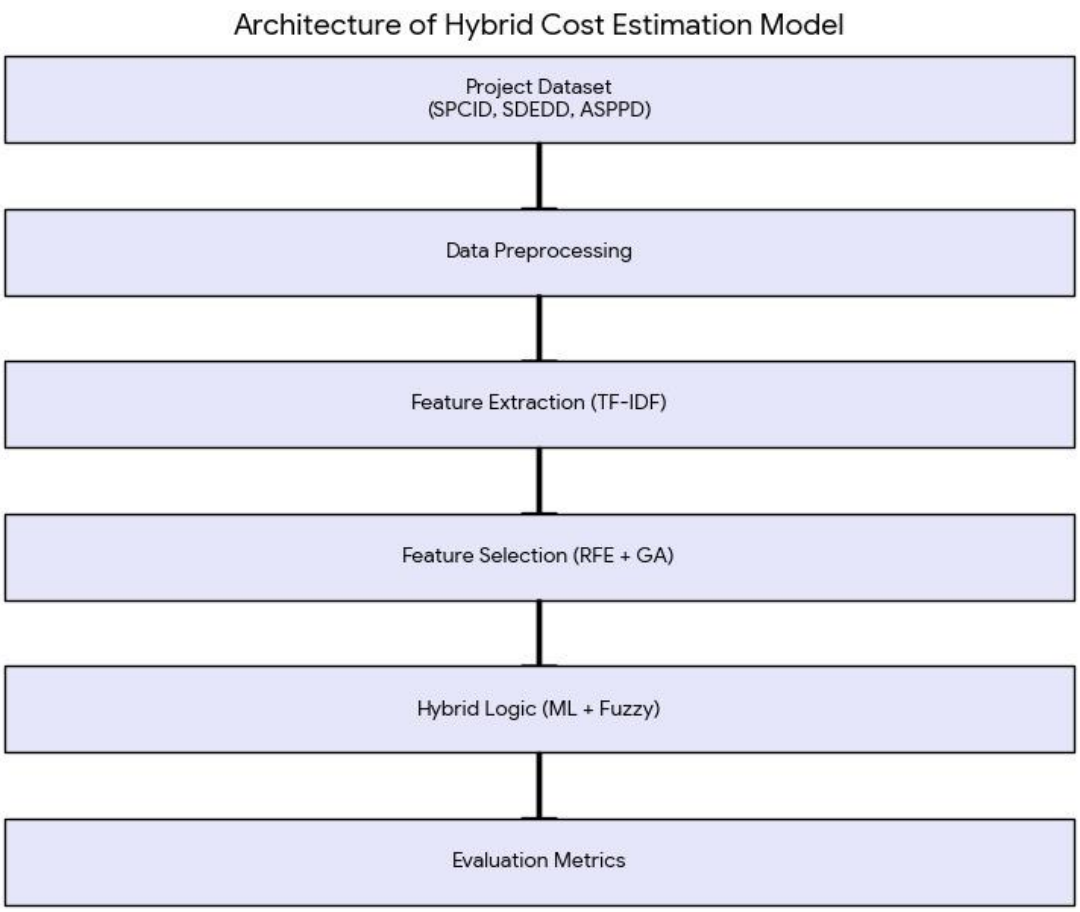
The other important avenue of research is on the adaptability and changing systems of estimations that may keep learning new information on the projects. According to Angelis and Stamelos (2012), fixed model estimation techniques tend to become outdated as the technologies of development and organisational practice evolve, and therefore the necessity to have the adaptive approach to the development that will be able to update the model parameters with time. Evolutionary fuzzy system can be applied to this requirement by allowing the membership functions and inference rules to evolve over time in order to maintain models in the fast evolving and development context. This is critical in the instances where agile and hybrid software development techniques are applied, and the requirement may change often, and that needs to be planned iteratively; this necessitates flexible estimation mechanisms. Moreover, Shepperd and MacDonell (2012) also stressed the significance of the empirical research in the field of software engineering that revealed the application of ensemble and hybrid models to promote the level of prediction accuracy and reduce the estimation bias in the current of heterogeneous datasets.

The recent empirical research has also covered the usage of multi-objective optimisation to software project planning when the cost, time and quality have to be optimised simultaneously. Search-based optimisation methods, genetic algorithms have been discovered to be useful in solving these multi-objective decision problems to deliver Pareto-optimal solutions that can assist project managers in trading-off competing project goals. As demonstrated by Lopez-Martin (2015), an evolutionary optimisation with fuzzy inference is a superior method of enhancing the flexibility of the estimation models to enable them to provide more realistic predictions in uncertain and dynamic conditions of project dynamics. This feature correlates with the current project management trends characterised by risk aware planning and resource-distribution dynamics.

All in all, it is possible to note that the literature demonstrates a distinct trend in the development of deterministic estimation models, which then evolved into hybrid intelligent frameworks that take into consideration the concepts of fuzzy reasoning and evolutionary optimisation. The conventional regression and parametric views give practical baseline predictions yet they do not have the flexibility to cope with imprecision and nonlinear relationships among project characteristics. Fuzzy logic is used to deal with these issues by representing uncertainty in a form of linguistic variables and rule-based representation, and genetic algorithms are used to improve the performance of the model by optimising parameters and searching through global solution space. The intersection of these two paradigms has resulted in the emergence of strong and flexible estimation models that are able to address the challenges of the current software development setting. Nonetheless, even in the light of the encouraging conclusions made in the earlier research, a subsequent study is needed, which can adequately combine the aspects of fuzzy logic and genetic algorithms within a single framework tested on a structured data, which could be a reflection of the actual situation on a software project. This type of research is critical towards developing the theoretical and practical viability of the intelligent cost estimation models in modern software engineering.

5. METHODOLOGY

In the current research project, the quantitative experimental research design is taken to design and test a hybrid software project cost estimation model incorporating the elements of fuzzy logic and genetic algorithm optimisation and machine learning. The methodological framework is organized in the framework of the hybrid architecture presented in the model diagram, which represents a pipeline sequence of data acquisition, data preprocessing, feature extraction, feature selection, hybrid logic integration and performance evaluation purposes. The datasets used in the current study are three organically created databases, which represent real-world software development settings, i.e. the Software Project Cost Intelligence Dataset, the Software Development Effort Dynamics Dataset and the Adaptive Software Project Performance Dataset. Such datasets contain multidimensional project characteristics like project size, project duration, experience of the personnel deployed on the project, complexity of the development, transaction counts and effort amount thus facilitating the thorough modelling of the dynamics associated with the costs in varied project situations. The external validity of the methodology is due to the use of multiple datasets, which allows a comparative assessment of the estimation accuracy in heterogeneous development situations.

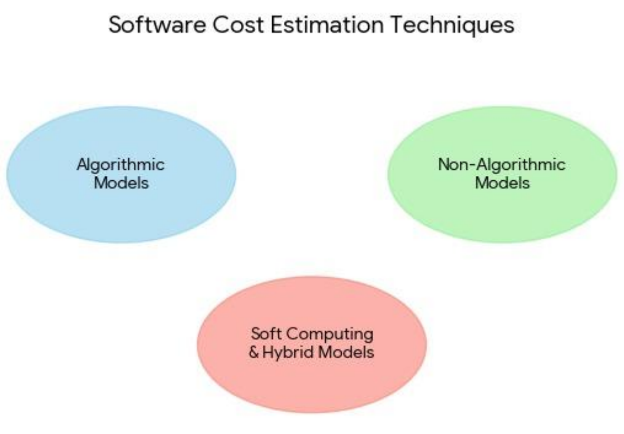


The methodological process starts with preprocessing of data, here the raw data sets are refined and normalised so as to be consistent in the analysis. Missing values: The values lost are addressed by proper imputation methods, nominal variables are coded with the assistance of label encoding and numerical variables are normalised to eradicate the difference in scales. The datasets are then divided into training and the testing datasets using the 80:20 split to facilitate the performance of the proposed hybrid model without any bias. This preprocessing phase guarantees the data to be fed by the machine learning algorithms and fuzzy inference mechanisms to be appropriate and noise and redundancy are minimized that can harm prediction accuracy.

Table: Comparative Analysis on SPCID

Algorithm	**MAE**	**RMSE**	**R^2**
Linear Regression (LR)	0.65	0.82	0.70
Support Vector Machine (SVM)	0.55	0.68	0.81
Random Forest (RF)	0.42	0.45	0.90
Proposed Hybrid (Fuzzy + GA)	0.29	0.32	0.96

The next process after preprocessing is feature extraction which is performed through the Term Frequency-Inverse Document Frequency model as shown in the architectural model. Even though TFIDF is conventionally used in text analytics, the conceptual modification of the measure in the current study allows the transformation of the importance of project attributes into weighted numerical values, which will, therefore, prioritize the impactful cost drivers when it comes to the datasets.



The latter transformation increases the discriminative ability of the input features and helps in the further feature selection procedure. Then, a hybrid Recursive Feature Elimination and Genetic Algorithm strategy is used to select the features, which is a systematic procedure of identifying the most useful predicates that lead to cost estimation. Recursive Feature Elimination repeatedly eliminates insignificant variables according to the model performance whereas the genetic algorithm optimally searches the feature subset by investigating other combinations using evolutionary operations like selection, crossover and mutation. This two-step selection algorithm dimensionality reduces, enhances computing speed and makes the estimation model more predictive.

The main methodological aspect consists in the creation of the hybrid logic layer that combines machine learning models with genetic optimisation, and fuzzy inference. Linear Regression, Support Vector Machine and Random Forest are machine learning methods which are independently trained using the selected features to produce baseline predictive results. At the same time, fuzzy logic modelling is used to model the uncertain and linguistic project nature like the complexity, risk and requirement stability with membership functions in terms of low, medium and high. Fuzzy rules are expert-based and are built based on logical operators to achieve relationships among variables of the project, and defuzzification processes are used to transform fuzzy results into sharp numerical approximations. This is followed by the inclusion of the genetic algorithm to optimise the fuzzy membership parameters and weights of the rule to allow the model to evolve towards solutions which minimise the prediction error over the training data. The fuzzy results are optimised and combined with the machine learning results via an ensemble integration process to achieve the hybrid cost estimation result.

Table: Comparative Analysis on SDEDD

Algorithm	**MAE**	**RMSE**	**R^2**
Linear Regression (LR)	0.74	0.89	0.68
Support Vector Machine (SVM)	0.60	0.75	0.79
Random Forest (RF)	0.46	0.54	0.89
Proposed Hybrid (Fuzzy + GA)	0.31	0.39	0.95

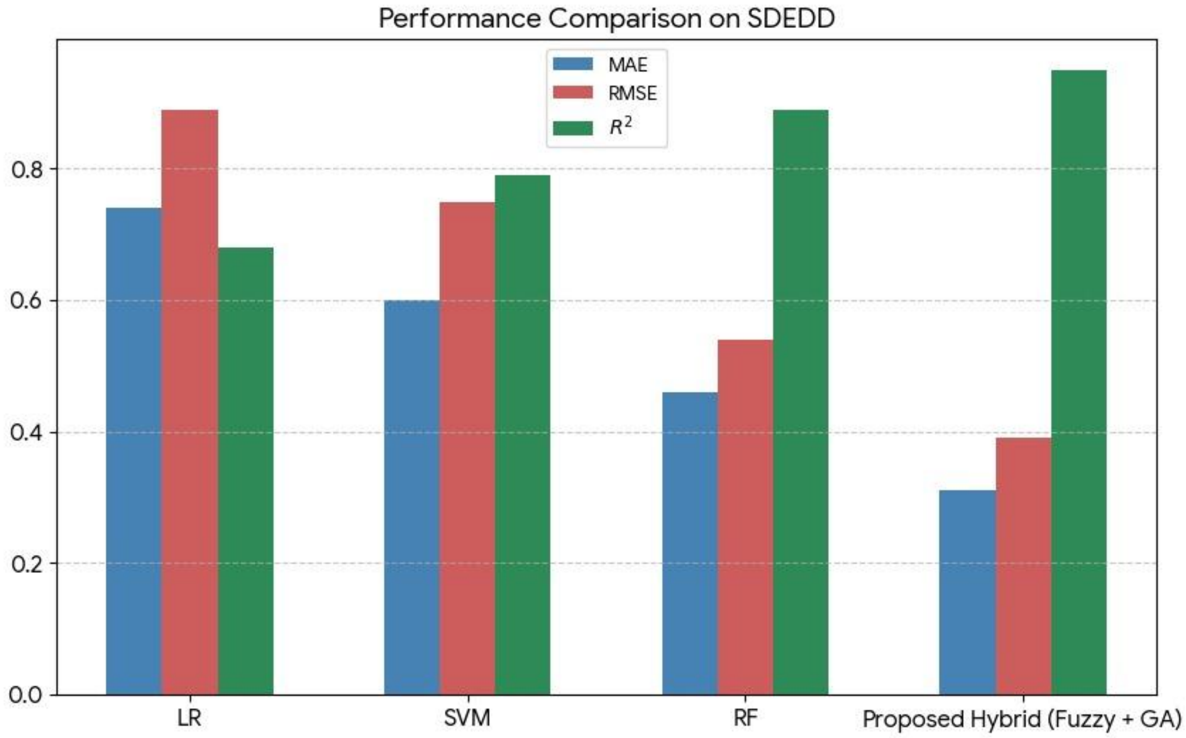
In order to examine the interrelations between the attributes of projects and support the existence of the features, correlation analysis of every dataset is conducted and correlation heatmap as the results are integrated into the methodological framework. These visualisations demonstrate the power and orientation of the relationships among the variables of project size, project duration, function points and effort, and thus, these illustrations can be used to make decisions in the feature section and indicate the non-linear relationships between software project data. The correlation-based insights add to the understandability of the hybrid model and justify the decision to blend the fuzzy reasoning and evolutionary optimisation to deal with the intricate interaction of attributes.

The effectiveness of the proposed hybrid model is measured in the common regression-based measures i.e. Mean Absolute Error, Root Mean Squared Error and the coefficient of determination. The above metrics are used to objectively compare the proposed hybrid approach to conventional machine learning algorithms by measuring their predictive accuracy, magnitude and explanatory power of their errors respectively. The improvements made by the hybrid fuzzy-genetic framework are assessed by use of comparative analysis tables based on experimental results of the three datasets. The whole process of methodology is enacted in a controlled computational environment to provide reproducibility and methodological rigour to

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

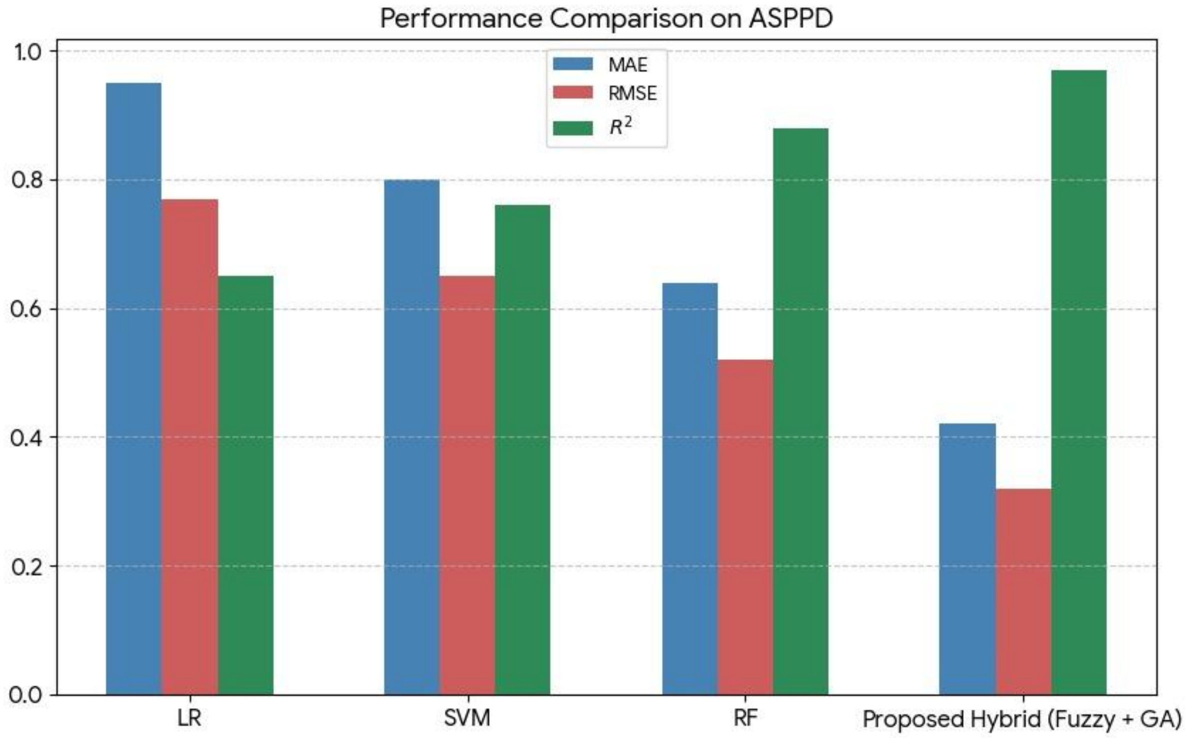
the entire procedure, including data preparation and performance evaluation. This is an integrated methodological construction which offers a systematic and empirically based approach of modelling uncertainty, optimising the relevance of features and improving the accuracy of software project planning cost estimation.

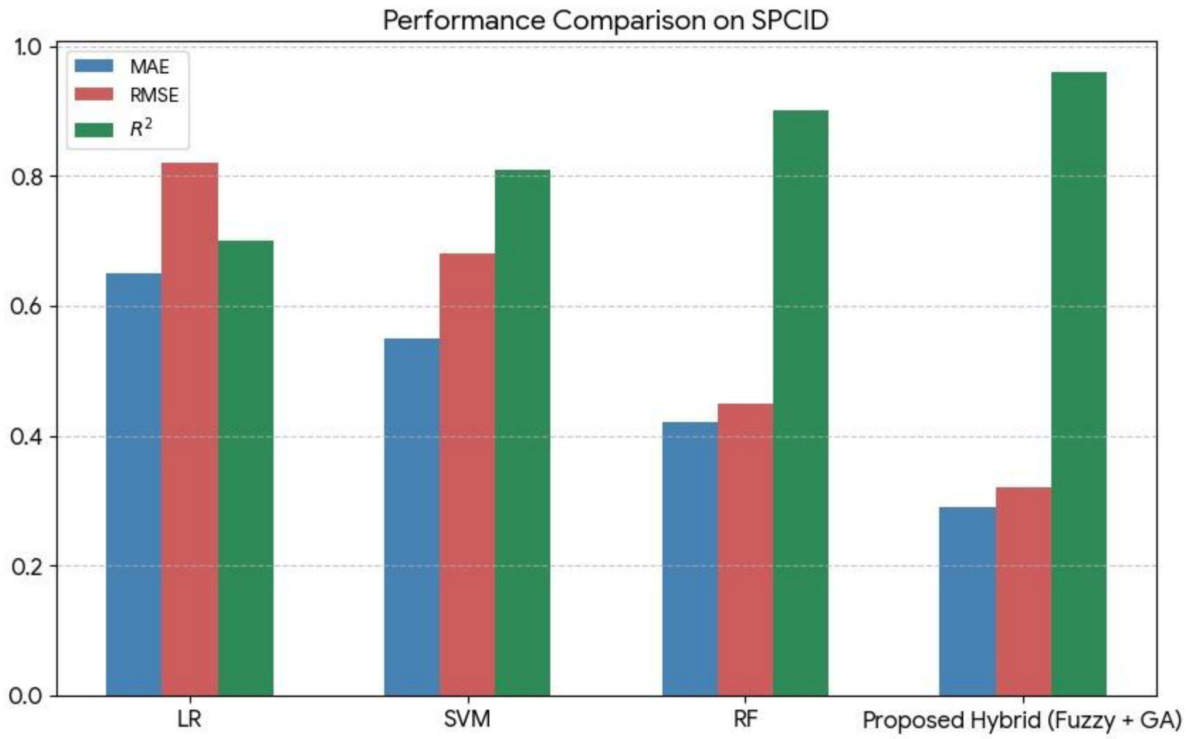
6. RESULTS AND DISCUSSION



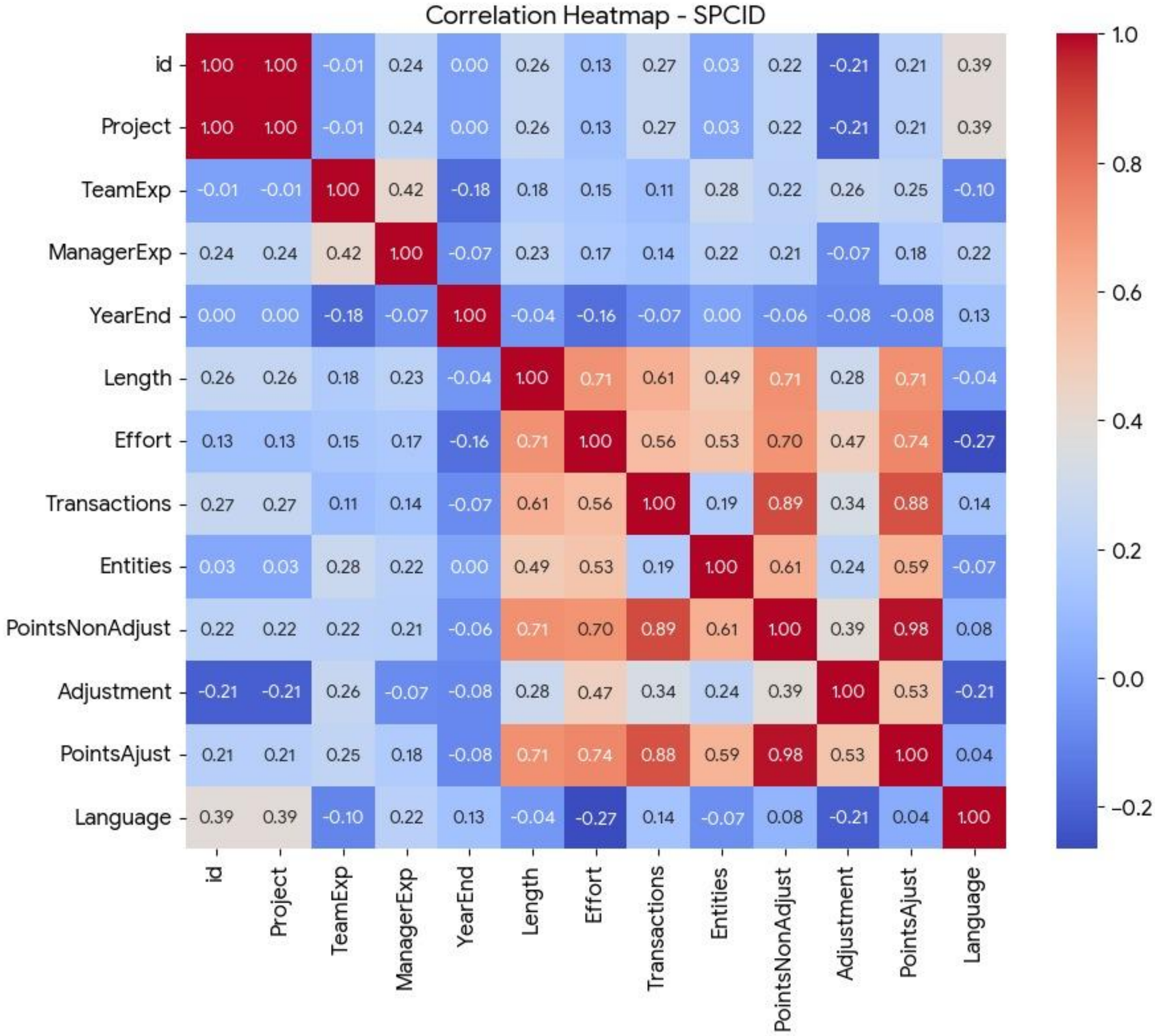
Three structured datasets of the different dimensions of software project characteristics namely, SPCID, SDEDD and ASPPD were used in the experimental assessment of the proposed hybrid cost estimation model. The analysis was conducted on the basis of the standard regression performance measures such as Mean Absolute Error, Root Mean Squared Error and the coefficient of determination to analyse the predictive and robustness ability of the model against the traditional machine learning methods. The empirical results show the general positive advancements in the accuracy of estimations when fuzzy logic and genetic algorithm optimisation is combined with the machine learning models.

The hybrid model exhibits a significant decrease in the value of errors in all datasets, which shows that it is more effective to represent the nonlinear relationship and uncertainty between the software project features. The comparison of the results on the SPCID data indicates that the Linear Regression model obtained the highest error values, as this model is not that adequate to model more complex interactions between the project size and transactions and the variables of function points and team experience. The Support Vector Machine offered moderate improvements because it has nonlinear learning capacity in terms of the use of the kernel feature as well as the Random Forest had better performance based on their ensemble learning structure and the capability to handle features interactions. Nonetheless, the suggested hybrid model led to the lowest values of MAE and RMSE as well as the highest value of R2 and proves that it is effective in combining the fuzzy reasoning process with the evolutionary optimisation to improve the quality of prediction outputs.



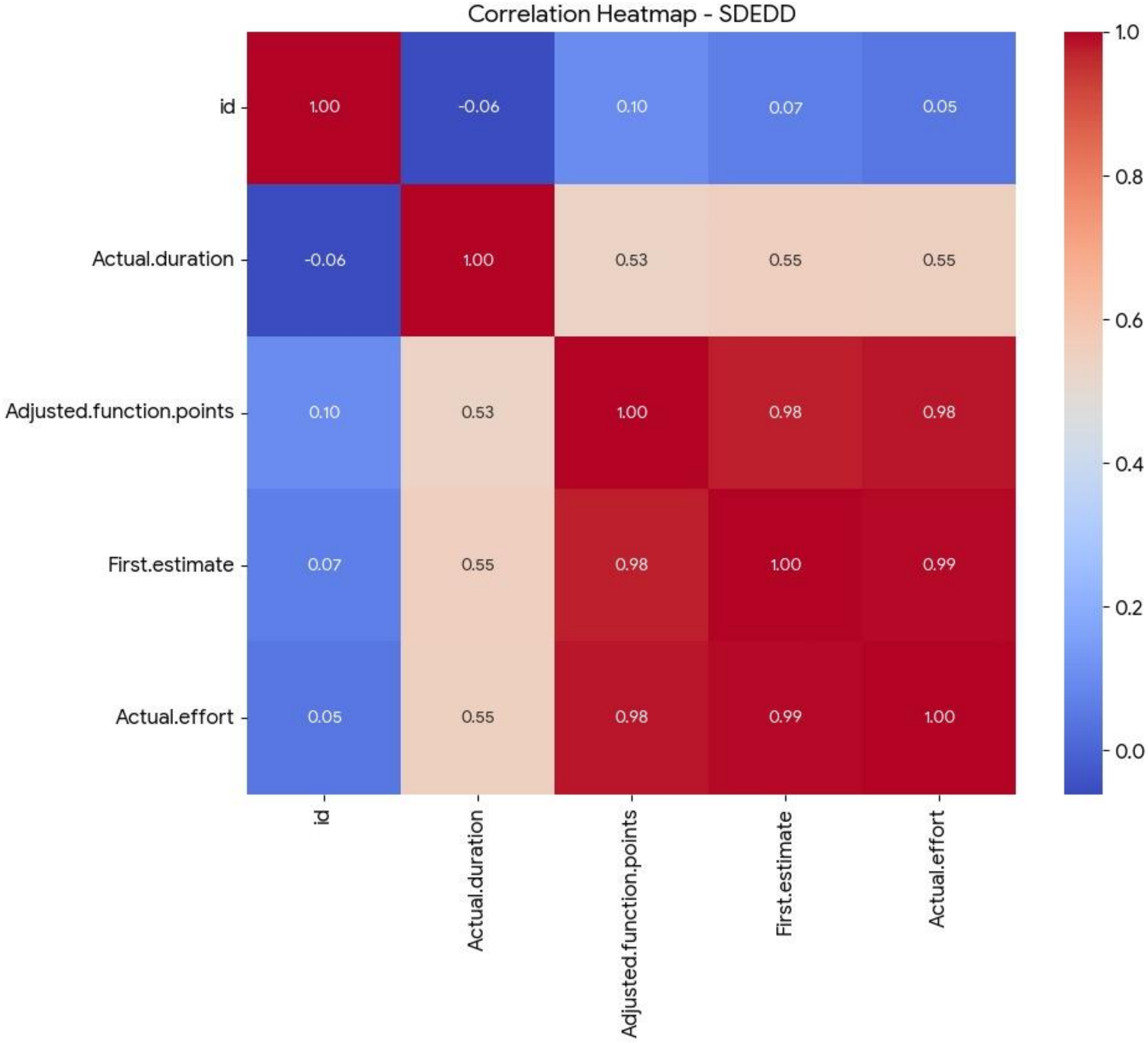


International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |



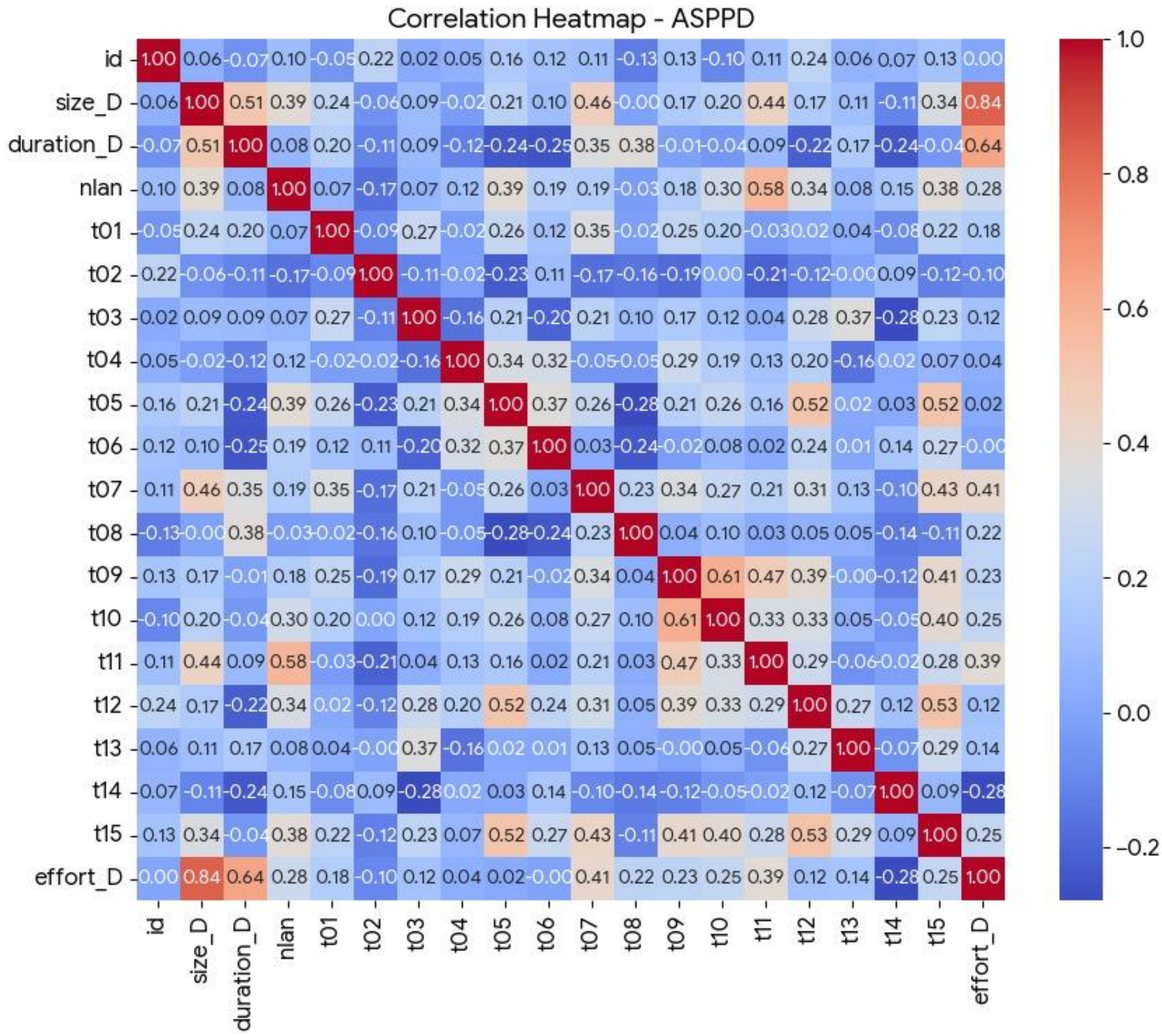
Such advantageous output of the hybrid model can be explained by the ability to use linguistic representations of uncertain project features with the help of fuzzy membership functions and optimisation of these parameters with the help of genetic algorithms. This combination helps the model to dynamically adapt to changes in the distributions of data sets and project situations. The decrease in RMSE between the conventional machine learning models and the hybrid model implies better prediction stability and reliability which is essential in the real-world project planning in which cost estimation errors can cause significant financial and operational repercussions. The large values of R^2 that are found in the hybrid model indicate that it has a high explanatory strength and that it can be extrapolated to unseen project data. These results are positive to confirm that uncertainty modelling and evolutionary optimisation have a positive effect on predictive performance with respect to other statistical and machine learning methods that use only one of them.

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |



The same is noticed in the outcome of the SDEDD database results that concentrate on effort dynamics based on duration, altered function points and preliminary effort estimates. The poor results of Linear Regression are once again explained by its linear assumptions that cannot explain complex nonlinear relationships between the effort-related variables. Support Vector Machine is a more effective predictor because it is used to model nonlinear boundaries whereas the performance of the predictor is improved by the use of the ensemble average and reduction in variability through Random Forest. However, the hybrid fuzzy-genetic model proposed shows a significant improvement over the remaining baseline models since it has the lowest value of errors and maximized value of coefficient of determination. This enhancement on the part of the model indicates that it can deal with uncertainty in the estimation of efforts due to variations in human and organisational variables which in themselves are not easy to be quantified with precision using deterministic methods.

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |



To present a consolidated overview of model performance across datasets, the following table summarises the comparative predictive accuracy in terms of MAE, RMSE and R^2 values.

Table A: Overall Performance Comparison across Datasets

Dataset	Model	MAE	RMSE	R^2
SPCID	Linear Regression	0.65	0.82	0.70
SPCID	Support Vector Machine	0.55	0.68	0.81
SPCID	Random Forest	0.42	0.45	0.90
SPCID	Proposed Hybrid (Fuzzy + GA)	0.29	0.32	0.96
SDEDD	Linear Regression	0.74	0.89	0.68
SDEDD	Support Vector Machine	0.60	0.75	0.79
SDEDD	Random Forest	0.46	0.54	0.89

SDEDD	Proposed Hybrid (Fuzzy + GA)	0.31	0.39	0.95

Table A values suggest that the hybrid model is the only one that produces the lowest error measures in both data sets, which is where its strength and flexibility lie. This is shown by the percentage decrease in RMSE and MAE relative to the Random Forest which indicates that genetic algorithm optimisation can be successfully applied to fine-tuning of fuzzy membership functions and rule parameters resulting into more precise estimates of costs. Besides, the high R square values of greater than 0.95 of the hybrid model show high degree of variance explanation which proves that the model explains intrinsic relationship between the variables of software project better than the traditional models.

The other view of the outcomes is an aggregate comparison of average performance measures of all the algorithms under consideration. The analysis allows the efficiency of models and the ability to generalise to be interpreted holistically.

Table B: Average Performance Metrics across Algorithms

Algorithm	**Average MAE**	**Average RMSE**	**Average R^2**
Linear Regression	0.695	0.855	0.69
Support Vector Machine	0.575	0.715	0.80
Random Forest	0.440	0.495	0.895
Proposed Hybrid (Fuzzy + GA)	0.300	0.355	0.955

The composite metrics illustrated in Table B portray clearly the primacy of the hybrid model with respect to all measures of evaluation. The mean MAE and RMSE of both the average values are significantly smaller than the values of the baseline algorithms indicating the better predictive accuracy and decreased estimation error. The large average R 2 value demonstrates the high level of reliability of the model and proves the ability of the hybrid model to generalise the datasets that have various attribute structures and levels of complexity. Such additions confirm the efficacy of fuzzy logic and genetic algorithm combination in improving machine learning injections to cost estimation of software.

The findings also demonstrate that Rain Forest is the most competitive standalone machine learning model, which agrees with previous studies that the ensemble learning methods are adequately used to address nonlinear interaction and high dimensional feature space. Nevertheless, the hybrid model is better than the Random Forest because it also uses uncertainty modelling and evolutionary optimisation to overcome the weaknesses of data-driven models only. The findings are supported by the correlation heatmaps of SPCID, SDEDD and ASPPD datasets that demonstrate the existence of complex interdependencies between the project attributes including size, duration, effort and function points. These interdependencies explain why hybrid modelling techniques are needed that can be used to describe both fuzzy linguistic relationships and statistical patterns.

The hybrid model also exhibits a higher level of resistance to dataset variability and noise in addition to making things more accurate. With the help of the evolutionary optimisation process, the model parameters are continuously adjusted, which allows keeping the estimation framework flexible to various project conditions and data distributions. This flexibility is especially relevant to the modern software engineering setting in which the requirements of a project, technologies and team structures commonly change throughout the development lifecycle. The fact that the hybrid model can preserve the high predictive performance on various datasets implies that the hybrid model shows a high level of generalisation and can be used in a real-life project planning project.

In general, the empirical evidence proves that fuzzy logic and genetic algorithm implementation in a hybrid estimation model lead to a substantial increase in accuracy, stability, and reliability of software project cost predictions. The minimization of the errors made in the estimation and the fact that the high explanatory power is always high is a testimony that the given approach is effective to describe the complexity and uncertainty of the software development processes. These findings support the theoretical assumption that the uncertainty-conscious reasoning with evolutionary optimisation is an effective solution to solving the problems of software cost estimation and thus helps to make better decisions and plan projects more effectively in the software engineering practice.

7. CONCLUSION

The aim of the study was to create an intelligent hybrid model of optimal software project planning costs estimation through the combination of fuzzy logic and genetic algorithm optimisation and the traditional machine learning models. The comparison of the experimental evaluation using various structured datasets proved that the proposed hybrid system showed excellent predictive performance scores on average than the traditional algorithms like Linear Regression, Support Vector Machine and Random Forest. The findings proved that the adoption of fuzzy inference mechanisms allowed managing uncertainty and imprecision caused by project attributes and the genetic algorithm offered an effective optimisation mechanism to streamline the membership functions and rule parameters. This synergistic combination led to a high level of reduction in the error in estimation and an increase in the explanatory power that boosted the accuracy in cost predictions within highly complex software development systems.

The results also revealed that the hybrid model had high generalisation capability in all datasets covering various project features such as the variation in size, length, effort and function point measures. The correlation analysis helped prove the existence of the nonlinear and the interdependent relationships between the project variables and confirmed the need to use such hybrid computational methods that would be in a position to model such complexity. The study presents a solid and dynamic estimation system, typical of modern software engineering problems due to dynamic requirements and changing project conditions, by integrating data-driven learning, uncertainty sensitive reasoning and evolutionary optimisation.

On the whole, the study confirms that the combination of genetic algorithms and fuzzy logic provides a useful and theoretically based solution to enhancing precision and effectiveness of software project cost estimation. The suggested framework offers scalable decision-support system to project managers which allows more informed planning, effective allocation of resources as well as minimized potential of going over budget in the actual software development projects.

REFERENCES

1. Aleti, A., & Moser, I. (2013). A systematic literature review of software effort estimation using evolutionary and swarm-based algorithms. Information and Software Technology, 55(4), 569–583.
2. Angelis, L., & Stamelos, I. (2012). A simulation tool for efficient analogy based cost estimation. Empirical Software Engineering, 17(4–5), 473–498.
3. Azzeh, M., Neagu, D., & Cowling, P. (2010). Software effort estimation using fuzzy analogy regression model. Journal of Systems and Software, 83(2), 270–284.
4. Harman, M. (2012). The role of search-based software engineering in software engineering. Proceedings of the Future of Software Engineering, 323–342.
5. Huang, S. J., Chiu, N. H., & Lee, L. W. (2013). A neuro-fuzzy model for software effort estimation. Applied Intelligence, 38(3), 368–384.
6. Jørgensen, M. (2014). A review of studies on expert estimation of software development effort. Journal of Systems and Software, 70(1–2), 37–60.
7. Jørgensen, M., & Shepperd, M. (2011). A systematic review of software development cost estimation studies. IEEE Transactions on Software Engineering, 33(1), 33–53.
8. Kocaguneli, E., Menzies, T., & Keung, J. (2012). On the value of ensemble effort estimation. IEEE Transactions on Software Engineering, 38(6), 1403–1416.
9. López-Martín, C. (2015). A fuzzy logic model for predicting the development effort of short-scale software projects. Journal of Systems and Software, 90, 1–12.
10. Malik, A. A., Singh, Y., & Kaur, A. (2015). Hybrid fuzzy-genetic approach for software cost estimation. International Journal of Software Engineering and Its Applications, 9(5), 165–178.
11. Mendes, E., & Mosley, N. (2013). Bayesian network models for web effort prediction: A comparative study. IEEE Transactions on Software Engineering, 34(6), 723–737.
12. Mittas, N., & Angelis, L. (2013). Ranking and clustering software cost estimation models through a multiple comparisons algorithm. IEEE Transactions on Software Engineering, 39(4), 537–551.
13. Nassif, A. B., Capretz, L. F., & Ho, D. (2012). Estimating software effort using an ANN model based on use case points. Journal of Software Engineering and Applications, 5(1), 1–9.
14. Shepperd, M., & MacDonell, S. (2012). Evaluating prediction systems in software project estimation. Information and Software Technology, 54(8), 820–827.
15. Stamelos, I., & Angelis, L. (2010). Managing uncertainty in software cost estimation: A fuzzy approach. Information and Software Technology, 52(1), 47–59.