REPUBLIC OF TÜRKİYE
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

PREDICTION OF ESTIMATE-AT-COMPLETION FOR CONSTRUCTION PROJECTS BY INTEGRATED ADOPTION OF ARTIFICIAL INTELLIGENCE, ARTIFICIAL NEURAL NETWORK AND OPTIMIZATION ALGORITHM

Ahmed ABOMAHDY

DOCTOR OF PHILOSOPHY THESIS

Department of Civil Engineering

Constrcution Engineering Program

Supervisor

Prof. Dr. Aslı Pelin Gürgün

Co-Supervisor

Assoc. Prof. Dr. Cenk BUDAYAN

January, 2025

REPUBLIC OF TÜRKİYE

YILDIZ TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

PREDICTION OF ESTIMATE-AT-COMPLETION FOR CONSTRUCTION PROJECTS BY INTEGRATED ADOPTION OF ARTIFICIAL INTELLIGENCE, ARTIFICIAL NEURAL NETWORK AND OPTIMIZATION ALGORITHM

A thesis submitted by Ahmed Ibrahim ABO MHADY in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY** is approved by the committee on 06.01.2025 in Department of Civil Engineering, Civil Engineering (English) Program.

Prof. Dr. Aslı Pelin GÜRGÜN Yıldız Technical University Supervisor	Assoc. Prof. Dr. Cenk BUDAYAN Middle East Technical University Co- supervisor

Approved By the Examining Committee

Prof. Dr. Aslı Pelin GÜRGÜN, Supervisor
Yildiz Technical University _____

Prof. Dr. Gül Polat TATAR, Member
Istanbul Technıcal Unıversıty _____

Prof. Dr. Pelin ÖZENER, Member
Yildiz Technical University _____

Asst. Prof. Dr. Kerim KOÇ, Member
Yildiz Technical University _____

Asst. Prof. Dr. Atilla DAMCI, Member
Istanbul Technıcal Unıversıty _____

I hereby declare that I have obtained the required legal permissions during data collection and exploitation procedures, that I have made the in-text citations and cited the references properly, that I haven't falsified and/or fabricated research data and results of the study and that I have abided by the principles of the scientific research and ethics during my Thesis Study under the title of Location Analysis of The Emergency Service Centers Of a Case Company supervised by my supervisor, Prof. Dr. Aslı Pelin GÜRGÜN and Co- supervisor: Assoc. Prof. Cenk BUDAYAN. In the case of a discovery of false statement, I am to acknowledge any legal consequence.

Ahmed ABO MHADY

Signature

This study was sponsored by T.C. Başbakanlık Yurtdışı Türkler ve Akraba Topluluklar Başkanlığı,

ACKNOWLEDGEMENT

To my first heroes, my mother and father: This accomplishment is as much yours as it is mine. I stand here because of your heartfelt prayers, unconditional love, and endless sacrifices.

To my parents-in-law, whose love, wisdom, and guidance have been a constant source of support throughout this journey—I am forever grateful for your encouragement.

To my little family: my dearest wife and my two beloved children. Your unwavering support, patience, and belief in me have been the cornerstone of this journey. Even a midst the war and challenges we faced, you have been my strength and inspiration.

To my family and friends, who inspire me to dream bigger every day: this work reflects your love and the gratitude I hold for all you have given me.

Finally, to the kind and wonderful people I met along this path, who brought joy and purpose to my life—may Allah bless your endeavors.

A scholar never progresses without guidance. Throughout my journey, I owe my deepest gratitude to my supervisor, Prof. Dr. Aslı Pelin GÜRGÜN and Co-supervisor: Assoc. Prof. Cenk BUDAYAN for their constant support and availability whenever I needed them. I am also sincerely thankful to the thesis monitoring committee, whose insights and encouragement inspired me to continue learning and striving for excellence.

TABLE OF CONTENTS

LIST OF SYMBOLS vii
LIST OF ABBREVIATIONS viii
LIST OF FIGURES xii
LIST OF TABLES xv
ABSTRACT xvi
ÖZET xviii
1 . INTRODUCTION 1
1.1 Background 1
1.2 Thesis Objective 4
1.3 Hypothesis 5
1.4 Concept of Estimation at Completion (EAC) 6
1.5 Research Scope and Limitations 8
1.6 Research Contribution 9
1.7 Outlines for Thesis 10
2 . LITERATURE REVIEW 11
2.1 Introduction 11
2.2 Definition of Cost Estimating 15
2.3 Purpose of Cost Estimation 16
2.4 Estimator Responsibility 19
2.5 Estimator Competences 20
2.6 Types of Estimates for Construction Costs 21
2.7 Indexed Formulation for EAC 22
2.8 EAC Computation Applying the Empirical Formula 24
2.9 Regression-Based Model for EAC Computation 27
2.10 Soft Computing Models for EAC 29
3 . METHODOLOGY 36
3.1 Overview of Adaptive Neuro-Fuzzy Inference System (ANFIS) 36
3.1.1 Membership Functions 37

3.1.2 Construction of ANFIS 38
3.2 Artificial Neural Network (ANN) 42
3.3 Variables Input Section 48
3.4 Optimization Algorithms: 49
3.5 Archimedes Optimization Algorithm: 49
3.6 FFA (FIREFLY) Optimizer Algorithm 52
3.7 Model Accuracy: 54
3.8 Modeling Procedure Phase 54
3.8.1 Programming Information and used Codes 54
3.8.2 The Work Methodology 60
3.9 Influence Variables 64
4 . RESULTS AND DISCUSSION **75**
4.1 Traditional Predictive Models 75
4.2 The Performance of Hybrid Predictive Models 77
5 . CONCLUSION **105**
REFERENCES **107**
PUBLICATIONS FROM THE THESIS **117**

LIST OF SYMBOLS

Y	The dependent variable (EAC)
β_0	The intercept term.
β_i	The coefficients of the independent variables X_i
ε	The error term.
Y_i	The actual value
$\bar{Y}_i$	The predicted value
$\bar{Y}$	The average of actual value.
$\bar{\bar{Y}}$	The average of predicted value.
N	The number of data

LIST OF ABBREVIATIONS

AACE	Association for The Advancement of Cost Engineering
AC	Actual Cost
ACp	Actual Cost Percentage
ACWP	Actual Cost Of Work Performed
AI	Artificial Intelligence
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
AOA	Archimedes Optimization Algorithm
BAC	Budget At Completion
BCWP	Budgeted Cost of Work Performed
BCWS	Budgeted Cost of Work Scheduled
BF	Brute Force
CCI	Construction Cost Index
CEAC	Cost Estimation At Completion
CEI	Climate Effect Index
CI	Composite Index
COI	Change Order Index
CPF	Construction Price Fluctuation
CPI	Cost Performance Index
CS	Cuckoo Search

CV	Cost Variance
DG	Distributed Generation
DNN	Deep Neural Network
DT	Decision Tree
EAC	Estimation At Completion
ECD	Estimated Completion Date
EFHNN	Evolutionary Fuzzy Hybrid Neural Network
EFSIM	Evolutionary Fuzzy Support Vector Machine Inference Model
ELM	Extreme Learning Machine
ESIM	Evolutionary Support Vector Machine Inference Model
ETC	Estimation To Completion
EV	Earned Value
EVM	Earned Value Management
EVMS	EVM System
EVp	Earned Value Percentage
FFA	Firefly Optimization Algorithm
FHNN	Fuzzy Hybrid Neural Network
FIS	Fuzzy Inference System
FL	Fuzzy Logic
FMGA	Fast Messy Genetic Algorithm
GA	Genetic algorithm
GHS	Global Harmony Search
GP	Genetic Programming
GPr	Gaussian Process

GUIS	Graphical User Interfaces
GWO	Grey Wolf Optimization
HNN	Hybrid Neural Network
HSA	Harmony Search Algorithm
IAOA	Integrated Archimedes Optimization Algorithm
IGA	Improved Genetic Algorithm
NN	Neural Networks
MAE	Mean Absolute Error
MPPT	Maximum Power Point Tracker
MRE	Mean Relative Error
MSE	Mean Square Error
OBI	Owner Billed Index
PBIAS	Percent Bias Errors
PSO	Particle Swarm Optimization Algorithm
PV	Planned Value
R	Correlation Coefficient
RMSE	Root Mean Squared Error
SA	Simulated Annealing
SBI	Subcontractor Billed Index
SCA	Sine Cosine Algorithm
SCI	Schedule Cost Index:
SMAPE	Symmetric Mean Absolute Percentage Error
SPI	Schedule Performance Index
SV	Schedule Variance

SVR	Support Vector Regression
SVM	Support Vector Machines
WSVM	Weighted Support Vector Machine

LIST OF FIGURES

Figure 2.1 The new intelligent model proposed aims to enhance the prediction accuracy of EAC [17]. 12

Figure 2.2 Construction Costs Classification.....18

Figure 3.1 Membership functions basic shapes [98]. 37

Figure 3.2 ANFIS framework[101]. 39

Figure 3.3 ANN architecture [105].....43

Figure 3.4 (a) The architecture of Feed Forward and (b) for Feed Back Neural Network topologies systems [106].....43

Figure 3.5*:* The AOA-ANN screenshot code. 56

Figure 3.6*:* The AOA-ANN screenshot code. 57

Figure 3.7*:* The AOA-ANFIS screenshot code.....58

Figure 3.8*:* The FFA-ANN screenshot code.....59

Figure 3.9: AFNIS and ANN models methodology 60

Figure 3.10: Hybrid AOA- ANFIS model Methodology. 61

Figure 3.11: Hybrid AOA- ANN model Methodology. 62

Figure 3.12: Hybrid FFA- ANN model Methodology.....63

Figure 3.13: Hybrid FFA- ANFIS model Methodology.....64

Figure 3.14 The developed EAC predictive model 68

Figure 4.1 The scatter plot diagram imagining between the actual and predicted values of EAC for traditional ANN model 76

Figure 4.2 The scatter plot diagram imagining between the actual and predicted values of EAC for traditional ANFIS model. 77

Figure 4.3 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 1 (FFA-ANN &FF-ANFIS). 83

Figure 4.4 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 2 (FFA-ANN &FF-ANFIS). 84

Figure 4.5 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 3 (FFA-ANN &FF-ANFIS). 85

Figure 4.6 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 4 (FFA-ANN &FF-ANFIS). 86

Figure 4.7 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 5 (FFA-ANN &FF-ANFIS). 87

Figure 4.8 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 6 (FFA-ANN &FF-ANFIS). 88

Figure 4.9 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 7 (FFA-ANN &FF-ANFIS). 89

Figure 4.10 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 8 (FFA-ANN &FF-ANFIS). 90

Figure 4.11 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 9 (FFA-ANN &FF-ANFIS). 91

Figure 4.12 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 1 (AOA-ANN &AOA-ANFIS).....92

Figure 4.13 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 2 (AOA-ANN &AOA-ANFIS).....93

Figure 4.14 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 3 (AOA-ANN &AOA-ANFIS).....94

Figure 4.15 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 4 (AOA-ANN &AOA-ANFIS).....95

Figure 4.16 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 5 (AOA-ANN &AOA-ANFIS).....96

Figure 4.17 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 6 (AOA-ANN &AOA-ANFIS).....97

Figure 4.18 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 7 (AOA-ANN &AOA-ANFIS).....98

Figure 4.19 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 8 (AOA-ANN &AOA-ANFIS).....99

Figure 4.20 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 9 (AOA-ANN &AOA-ANFIS).....100

Figure 4.21(a): Taylor diagram for FFA-ANN Prediction Models.....101

Figure 4.22(b): Taylor diagram for FFA-ANFIS Prediction Models.....102

Figure 4.23 (a): Taylor diagram for AOA-ANN Prediction Models.....102

Figure 4.24 (b): Taylor diagram for AOA-ANFIS Prediction Models 103

Figure 4.25: The Spyder diagram for ANN, ANFIS, and best models in AOA-ANN & AOA-ANFIS.....104

Figure 4.26: The Spyder diagram for ANN, ANFIS, and best models in AOA-ANN & AOA-ANFIS.....104

LIST OF TABLES

Table 2.1 EAC Comparisons results (13 Production and 14 Development Contracts (Navy)) 25

Table 2.2 Comparison of EAC Results from 11 Development Contracts Administered by the Army 26

Table 2.3 EAC Comparisons Results (40 Production Contracts and 16 Development (Air Force) 27

Table 3.1 Advantages and disadvantages of the ANN, ANFIS, and hybrid AOA/ FFA -ANN/ANFIS 46

Table 3.2 Information of historical data …. 65

Table 3.3 Factors that exert an impact. 66

Table 4.1 The numerical performance metrics for the ANFIS and ANN predictive models, referred to as the "Base-model versions," during the testing phase of modeling. 76

Table 4.2 The feature combination elements to predict EAC value using FFA-ANN model 78

Table 4.3 The FFA-ANN prediction model's numerical indicators for the deferent combination modeling phase 78

Table 4.4 The feature combination elements used to predict EAC value using FFA-ANFIS model 79

Table 4.5 The FFA-ANFIS prediction model's numerical indicators for the deferent combination modeling phase 79

Table 4.6 The feature combination elements used to predict EAC value using AOA-ANFIS model 80

Table 4.7 The AOA-ANFIS prediction model's numerical indicators for the deferent combination modeling phase 80

Table 4.8 The feature combination elements used to predict EAC value using AOA-ANN model 81

Table 4.9 The AOA-ANN prediction model's numerical indicators for the deferent combination modeling phase 81

ABSTRACT

Prediction of Estimate-at-Completion for Construction Projects by Integrated Adoption of Artificial Intelligence, Artificial Neural Network and Optimization Algorithm

Ahmed ABO MHADY

Department of Civil Engineering
Doctor of Philosophy Thesis

Supervisor: Prof. Dr. Aslı Pelin Gürgün
Co-supervisor: Assoc. Prof. Cenk BUDAYAN

In the construction project management industry, determining the final project cost involves considering various influencing factors. Among the different methods, Estimate at Completion (EAC) stands out as a crucial approach for projecting the project's final cost. The key advantage of EAC lies in its ability to incorporate both the likelihood of project performance and associated risks. Moreover, EAC proves highly beneficial for project managers as it helps identify critical issues throughout the project's progression and enables the determination of appropriate solutions. This study introduces innovative coupled intelligence models, specifically integrating the Firefly Optimization Algorithm (FFA) and the Archimedes Optimization Algorithm (AOA) with Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and Artificial Neural Network (ANN) for modeling project construction estimates at completion. AOA and FFA are utilized to identify significant attributes affecting the EAC which is dependent variable. This approach underscores the efficacy of hybrid models as a novel predictive framework within this domain. A traditional ANN is designed as a benchmark to evaluate the predictive accuracy of the hybrid models. These models are built using historical data from construction

projects performed in Taiwan covering the period from 2000 to 2007. The study aims to determine EAC and assess the trend change in the forecast model monitor. The primary objective is to establish a reliable trend of EAC estimates, aiding the construction companies in enhancing project cost control effectiveness. The results highlight that the FFA-ANN model shows superior performance over traditional ANN, ANFIS models and hybridized AOA-ANN, AOA-ANFIS and FFA-ANFIS models.

Keywords: Neural Network; Cost estimation; Archimedes Optimization Algorithm (AOA), Firefly optimization algorithm (FFA), Adaptive Neuro-Fuzzy Inference Systems (ANFIS).

ÖZET

Yapay Zeka, Yapay Sinir Ağı ve Optimizasyon Algoritmasının Entegre Kullanımı ile İnşaat Projeleri için Projenin Tamamlanmada tahminlenen maliyetinin Öngörülmesi

Ahmed ABO MHADY

İnşaat Mühendisliği Bölümü

Doktora Tezi

Danışman: Prof. Dr.Aslı Pelin Gürgün

Eş-Danışman: Assoc. Prof. Cenk BUDAYAN

Yapım Proje Yönetimi sektöründe, nihai proje maliyetini belirlemek çeşitli etkenlerin dikkate alınmasını gerektirir. Farklı yöntemler arasında, Proje Tamamlama Tahmini (EAC) proje maliyetinin son durumunu öngörmede kritik bir yaklaşım olarak öne çıkmaktadır. EAC'nin temel avantajı hem projenin başarı olasılığını hem de ilişkili riskleri dikkate alma yeteneğidir. Ayrıca, EAC proje yöneticileri için oldukça faydalıdır, çünkü projenin ilerlemesi boyunca EAC hem proje yöneticilerinin kritik sorunları ve hem de bu sorunlara uygun çözümleri belirlemesine yardımcı olur. Bu çalışma, EAC'yi modellemek için Ateşböceği Optimizasyon Algoritması (FFA) ve Arşimet Optimizasyon Algoritması (AOA) ile Adaptif Ağ Yapısına Dayalı Bulanık Çıkarım Sistemi (ANFIS) ve Yapay Sinir Ağı (ANN) birleştirerek dört yenilikçi birleşik zekâ modeli önermektedir. AOA ve FFA,

EAC bağımlı değişkenini etkileyen etkili özellikleri çıkarmak için kullanılmaktadır. Geliştirilen modellerin EAC tahmin performanslarını ortaya çıkarmak için aynı zamanda başka tahmin modelleri geleneksel ANN ve ANFIS ile geliştirilmiştir. Daha sonra bu modeller ile geliştirilen hibrit modeller karşılaştırılarak geliştirilen modellerin EAC tahminlerinin doğruluğu test edilmiştir. Tahmin modelleri, 2000 ile 2007 yılları arasında Tayvan'daki inşaat projelerinden alınan tarihsel bilgiler kullanılarak geliştirilmiştir. Çalışmanın amacı, EAC'yi belirlemek ve tahmin model monitöründeki eğilim değişikliğini değerlendirmektir. Birincil amaç ise proje yöneticilerine proje maliyet kontrol etkinliğini artırmada yardımcı olacak güvenilir bir EAC tahmin eğilimi oluşturmaktır. Sonuçlar, FFA-ANN modelinin klasik ANN, AOA-ANN ve geleneksel ve hibrit ANFIS modellerine göre önemli ölçüde üstün olduğunu vurgulamaktadır.

Anahtar Kelimeler: Sinir Ağı; Maliyet Tahmini; Arşimet Optimizasyon Algoritması (AOA); Ateşböceği Optimizasyon Algoritması (FFA); Uyarlanabilir Nöro-Bulanık Çıkarım Sistemleri (ANFIS).

1 INTRODUCTION

1.1 Background

As the management of construction projects evolves, so do the challenges associated with the industry. This result is largely due to the unique and ever-changing nature of the construction industry. For instance, even a small size project involves many stakeholders. The involvement of contractors, stakeholders, clients, investors, consultants, and regulators makes construction projects inherently complex and stressful. Unlike other industries where selling prices are primarily based on established production costs, the construction industry needs to estimate project costs through feasibility studies, engineering design, cost budgeting, and management before construction begins [1]. Consequently, construction project management is characterized by its inherent risk and complexity.

As Zeng et al. [2] pointed out, the construction industry is inherently dynamic and characterized by rapid and continuous changes. Time is often limited, and there is pressure to save on construction costs while maintaining top-notch quality standards. All these factors make the management of construction projects highly demanding and challenge their successful completion. This often leads to poor project performance and budget overruns. Therefore, it is critical to maintain an up-to-date plan that accurately keeps up with the project's current conditions. Using an approach where timely intervention is possible enables construction projects to be managed effectively and allows for making adjustments when the project appears to be going off track. Overlooking the requirement of employing such an approach may lead to catastrophic outcomes, such as cost overruns detected in the late phases of projects, making it too late to respond and adapt with corrective measures.

Successful Project management requires utilizing expertise, tools, skills and appropriate management approaches to projects while taking into account all

projects' limitations and constraints. For a timely and successful completion of a construction project, time, cost and quality are important constraints that have be accounted for [3]. A crucial factor that can significantly impact these constraints is the team-rounded focus. These constraints are known for their interconnectedness. This means that deviations in any single constraint can shift the burden onto others [4]. For example, when a there is a delay in a project's completion, the project's costs will rise to make up for the resources that need to be allocated in order to minimize the time delay, and this can lead to overall extra costs to the projects budget [5]. Alternatively, the companies can prefer delivering a product with lower quality standards to ensure the timely delivery of the project's final product within a budget [6]. Thus, changes in time, cost, and quality parameters can bring additional risks to the project, emphasizing the importance of good change management throughout its lifecycle.

Subsequently, it became evident that a successful completion of a project requires an in-depth evaluation of the surrounding circumstances, coordination between constraints and the establishment of robust stakeholder relationships. As a result, it is crucial to consistently assess the progress of a project based on available information in order to have a timely and successful project completion. As the project advances, the information becomes more defined, allowing for more precise estimations that facilitate successful project completion.

Although effective cost management is often time-consuming and complex, due to the numerous factors affecting project costs, that must be addressed separately and adequately at the various stages of a project [7], it is vital to execute the processes seamlessly and timely in construction projects. In particular, the significance of applying a proactive cost control system is frequently emphasized to identify potential issues in cost management. In construction projects, cost control, which mostly relies on the expert judgment, should be performed using scientific methods to improve accuracy. Consequently, the Estimate at Completion (EAC) was proposed for cost controlling and has been widely used in construction projects [8,9], as it enables project managers to foresee potential issues and plan appropriate responses. However, forecasting the EAC is a challenging task due to the considerable uncertainty created by the large amount of variable data characterizing construction costs [10].

During the planning stage, the managers of the project intended to keep the project's budget under control at beginning of the process and ignore the cost adjustments impact and updates of information during subsequent project stages [11]. This results in ineffective control of cost and the inability to forecast the coming difficulties. Also, the changing environment of the building sector repeatedly results in budget adjustments, which are viewed as a vital part of cost management. The precise estimation of costs is paramount to the triumph of a project, as it empowers managers to adapt their tactics and decisions to manage potential risks. The primary approach utilized for this aim is earned value management (EVM), a systematic structure for assessing and measuring a project's efficiency and the forecasted expenses and timeline at completion. The EVM methodology relies on project expenditures, including actual cost (AC), earned value (EV), and planned value (PV), to monitor the advancement and performance of activities, making it a versatile tool that can be employed for any project. Moreover, estimation at completion (EAC) is frequently calculated by project managers using EVM, which is an effortless and automated method for computing the projected costs of completing predetermined tasks. Management can assess the variance between planned and actual expenses using the projected EAC and identify the underlying issue. In practice, managers collect cost management data first and then calculate the EAC using formulas [12]. One disadvantage of utilizing formulas is that a variety of techniques for calculating the EAC are possibly available. At least eight techniques have been identified in prior study publications. Thus, project managers are required to exercise their discretion in selecting the optimal approach for computation. Needless to say, the factors that influence project costs are manifold, and each project possesses its own distinct characteristics. This explains why a given formula may be suitable in some cases but may be unsuitable in others [9]. (EV) can assist project managers in being aware of the program’s status, both pertaining to performance and cost categories, if they are used and analyzed correctly. In order to estimate and forecast future status, the EVM system (EVMS) is used for organizational project scheduling, budgeting, and component planning. Program managers can predict the outcomes of a decision-making process by utilizing the EVMS. The accuracy of expected estimations is what affects a decision-making process's efficiency. The EVMS approach offers the inputs (referred to as Estimates at Completion (EAC)) for these forecasts. Certain

unfavorable program outcomes (such as cost overruns, cancellations, and timetable delays) are the result of an inefficient decision-making process caused by faulty estimations [13,14].

Several studies have been developed to address the EAC's related problems, these studies incorporated numerous input variables (including but not limited to cost variance (CV), schedule variance (SV), subcontractor billed index (SBI), schedule performance index (SPI), cost performance index (CPI), climate effect index, change order index (COI), owner billed index, and construction price fluctuation (CPF)) to simulate the EAC. To establish the relationship using artificial intelligence (AI) Various AI models, such as the adaptive neuro-fuzzy inference system (ANFIS), ANN, support vector regression (SVR) and genetic programming (GP), and many others have been devised to address issues related to the EAC. However, because of the high complexity of the problem related to the nonlinearity, non-stationarity, and stochasticity of the relationships between the independent variables, various limitations were identified in the literature using the proposed AI models. Several AI disadvantages include the demand for internal parameter optimization, trapping in local minimal difficulties, and the requirement for human interaction during the neural network structure construction process. Thus, to the best of our understanding, a novel hybrid intelligent model was employed in this research to calculate the EAC. The model integrates ANFIS and ANN with the firefly optimization (FFA) algorithm and Archimedes optimization algorithm (AOA). This study was primarily concerned with identifying the most effective EAC approaches for the previously listed moderator variables. There was no attempt made to deploy the proposed predictive model as an actual expert system on a practical assignment.

1.2 Thesis Objective

As determining the EAC is considered challenging, it is essential to develop a swift and efficient system that incorporates cost control during project implementation to anticipate project EAC using AI techniques. This research aims to improve and resolve the effective issues associated with project cost management by using real-world historical data and conducting investigations into project cost management in order to identify the effective variables affecting project cost. Given the limits of

present Artificial intelligence models, it is extremely promoting to investigate more dependable, resilient, and trustworthy approaches for resolving cost overruns throughout project implementation. From many construction projects, Historical data were analyzed to determine the suggested model's predictability. By utilizing project data, a pattern for the cost flow of the project, as well as the relationship between the project's EAC and monthly expenses, was established based on previous experience and knowledge. The research objectives can be summarized as follows:

i. A novel approach was adopted to model the EAC by developing hybrid intelligent models. In these models, the ANFIS and ANN framework were integrated with the firefly and Archimedes optimization algorithms.

ii. The suggested model's predictability performance in computing EAC was compared to that of a traditional ANN, which is one of the most often used AI model for EAC estimation.

iii. Input attribute optimization approaches (firefly and Archimedes optimization algorithms) was used to identify of the features that significantly affect project cost. This lead to enhance the predictability of the ANFIS and ANN prediction models.

iv. In summary, the research examined a novel modeling technique according to the linked intelligence model that could support project managers to make the decision.

1.3 Hypothesis

The ability to prepare and evaluate the worst-case scenario, best-case scenario, and EAC are gaining the attention of industrial practices. Each year, a comprehensive EAC based on the "bottoms-up" approach is mandatory to comply with the EVMS requirements [15]. Furthermore, an initial EAC is often formulated during the initiation phase of a project or construction to reduce the degree of uncertainty associated with design and/or material costs. This aids project managers in addressing the following queries:

- Will the residual funds be sufficient to finish the project?
- Could previous cost experiences be used to predict future cost outcomes?
- Can the remaining project be altered in light of its performance to date?
- Can project cost efficiency include an effect on the financial health of the company?

Thus, a realistic and timely Estimated Completion Date (ECD) and EAC can be critical components of corporate finance and project management applications. Additionally, both require routine differentiation of the ECD and EAC in addition to the engaging earmarks in order to ensure a financially viable performance for investors and clients.

1.4 Concept of Estimation at Completion (EAC)

Project managers are often under pressure from organizations and clients to ensure that their projects are completed within the approved budget, the budget at completion (BAC), and the planned completion date. The EAC and ECD are important metrics that indicate progress in addressing these concerns. According to the EV standards, the EAC is the sum of the actual cost of work performed (ACWP) under the contracts and the best estimate of the project manager for the time-phased resources required to accomplish the remaining approved work (ETC). This relationship is frequently expressed as follows:

$$EAC = ACWP + ETC \tag{1.1}$$

Therefore, the EAC can be regarded as a projection of the ultimate cost of the project. In order to ensure that the project's goals are achieved within the allocated budget, the project manager may need to reevaluate task priorities and reschedule or re-plan the remaining project activities. The main aim is to complete the project within budget and within the stipulated timeline.

The degree of uncertainty associated with remaining tasks, the available knowledge, and the perceived risks may vary in all estimates. Therefore, responsible project management necessitates determining the EACs' validity, particularly when they vary significantly from the approved project budget. Consequently, project management's primary concern is to assess the level of uncertainty connected with the remaining tasks, estimate the cost of the remaining tasks, and manage the impact of the uncertainty on the project's cost objectives to ensure successful project completion within budget and time constraints.

Therefore, the Contract Performance Report and the Integrated Program Management Report necessitate three separate EACs during bidding to gather information on the extent of cost uncertainty and unforeseen project risks. These

EACs are required to represent the worst-case scenario, the best-case scenario, and the most probable scenario, which is the project manager's best estimate, in order to forecast the ECD.

When calculating the EAC, the uncertainty is influenced by the EAC itself since the actual cost-to-date is already known. The ETC entails reassessing the resources needed to complete the remaining approved work, relying on past cost experience, and considering several factors like direct and overhead costs. Furthermore, an accurately estimated ETC must consider various factors such as expected labor rates, labor efficiency, purchase order commitments, material prices and consumption, prices and usage performance of other direct costs, resource types, risk and opportunity, and other relevant factors as identified by senior management. It must also be aligned with the current schedule and be in accordance with the ECD.

Commonly, EAC estimation involves mathematical computations or an autonomous examination of cost and schedule information. A method employed in this process entails utilizing performance metrics like the CPI and the cumulative budgeted cost for work performed (ACWP) to ascertain the EAC. This is accomplished by dividing the project's BAC by the CPI.

The computation-based EAC is often called an Independent EAC. It is different from the formal EAC because it can be quickly generated and used to verify the reliability of the existing cost estimate and to determine if a comprehensive EAC is needed. It is important to note that such calculations do not take into account expected labor rates and efficiency, risk and opportunity, schedule risk assessment (SRA), and other factors. Despite claims of being free from subjectivity and bias, the computation of EACs often involves comparison analysis and requires judgment and logical reasoning.

Prior to commencing a significant project, a detailed EAC is typically prepared, taking into account the decreased uncertainty resulting from the publication of project design and bill of materials. The comprehensive EAC helps the project manager address the following inquiries:

- Can the project be completed within the approved budget?

- Are there any adjustments that can be made to the remaining project tasks based on the current performance?
- How can the prior cost experience help in forecasting future cost performance?
- What is the impact of the project's cost performance on the overall financial health of the organization?

Therefore, it is crucial to release the EAC and ECD on time, as they play a vital role in both project management and financial management. Continuous monitoring and comparison of these indicators against project targets are necessary to accurately predict financial performance for investors and clients. This step is critical in building trust and maintaining a positive reputation in the industry [16].

1.5 Research Scope and Limitations

This study aims at the use of a hybrid intelligent model to determine the complex nonlinearity/non-stationarity/stochasticity relationship between the related independent variables (i.e., Construction progress percentage, Actual cost percentage (ACp), Earned value percentage (EVp), CPI, SPI, SBI, Owner billed index (OBI), COI, Construction cost index (CCI) and Climate effect index (CEI)) and the dependent variable (i.e., EAC). The existing literature showed that EAC had been calculated using several versions of artificial intelligence (AI) models, such as an ANN, ANFIS, support vector machine (SVM), GP and decision tree (DT). However, several limitations indicated over the literature based on the developed AI models. Among several AI drawbacks, the requirement for the internal parameters' optimization, trapping in the local minima problems, and the human interaction for constructing the neural network structure. Hence and to the best knowledge of the current study, a novel approach was adopted to model the cost estimation at completion (CEAC) by developing a hybrid intelligent model that integrates the ANFIS and ANN framework with the firefly and Archimedes algorithms. The project faces limitations primarily rooted in the challenges of gathering data from construction projects, compounded by security and political issues and to overcome these limitations, the historical data used for the analysis was taken from [17,18,19] studies. The data was believed to provide a sufficient representative of the studied database. This study mainly focused on the

identification of the best performing EAC methods concerning the previously discussed moderator variables. There was no attempt on the implementation of the proposed predictive model on a practical project as an actual expert system.

1.6 Research Contribution

EAC prediction is of paramount importance in construction management, synergizing with the critical aspect of accurate cost prediction. EAC serves as a pivotal tool to ensure that a construction project adheres to its budget and timeline. While initial cost estimates lay the groundwork, ongoing EAC predictions empower stakeholders to adapt and make informed decisions as the project unfolds. The inherent dynamism of construction projects, coupled with potential uncertainties, underscores the need for a robust EAC prediction methodology. Stakeholders, including owners, contractors, and planners, rely on EAC forecasts to proactively manage expenses, optimize resource allocation, and assess the project's financial feasibility throughout its lifecycle. Moreover, EAC predictions contribute to strategic decision-making by providing insights into potential cost overruns or savings, enabling proactive measures [20]. This research significantly contributes to construction management by introducing a groundbreaking approach to modeling EAC. The innovative methodology integrates the ANFIS and ANN frameworks with the firefly and Archimedes algorithms. This amalgamation of advanced computational techniques aims to enhance the accuracy and reliability of EAC predictions in construction building projects. By harnessing the adaptive and learning capabilities of ANFIS and ANN, along with the optimization power of firefly and Archimedes algorithms, the proposed model offers a sophisticated and robust solution to address the dynamic and complex nature of construction projects. This integration is poised to result in a comprehensive EAC prediction tool, empowering construction professionals to proactively manage costs and resources throughout the project lifecycle and paving the way for further advancements in predictive analytics within the construction management domain.

1.7 Outlines for Thesis

The following sections address the study's objectives.

The opening section of this doctoral dissertation presents a comprehensive overview of the research, specifying the reasons, objectives, and organization of the project.

In the second chapter of this research, an extensive review of the existing literature on estimating at completion for large engineering projects is carried out. The chapter begins with a detailed analysis of traditional EAC computations. It then provides an up-to-date review of research conducted in the last 20 years, covering the latest advancements in artificial intelligence and complementary models. The chapter concludes with an evaluation of the studies reviewed.

In the third chapter, the ANFIS and ANN is presented as a proposed method for estimating project completion. The chapter also includes a discussion on the optimization of inputs by using Firefly and Archimedes optimization algorithms and provides a brief explanation of the dataset used in the case study. The process of creating the hybrid models is also described in detail.

In Chapter four, a detailed analysis of the outcomes of the proposed models was conducted. The effectiveness of the models was evaluated through various performance measures. The chapter includes a comprehensive discussion and analysis of the application, including the efficiency of the models, the performance of individual and combined models, and the ability of the proposed method to capture the nonlinearity of the problem under investigation.

In conclusion, Chapter five presents a summary of the research's findings and contributions to the field of project management modeling. The chapter also outlines several suggestions for future research, with a specific focus on estimating at completion.

2
LITERATURE REVIEW

2.1 Introduction

In this chapter, an extensive literature review on EAC is presented, covering its origins, developments, constraints, and usage in different projects. The chapter starts with a brief history of EAC, followed by an overview of its extensions and variations over time. Additionally, the limitations of the EAC method are analyzed in detail. Finally, the chapter examines algorithm applications in different projects. Additionally, it includes a survey of the literature regarding the parameters used to create single and hybrid models, as well as their application in different projects.

The construction industry often suffers from poor performance due to the high level of risk associated with project management. The volatile nature of the industry, combined with external constraints, has made risk management a significant challenge [2,21]. To create a profit, the monitoring of a project should be provided proper attention (in terms of taking suitable corrective action when deviations occur, as well as close monitoring and identification of deviations). Most construction projects typically prioritize budget planning during the initial stage, without considering the potential impact of changes in engineering costs or updated information throughout the construction process [22], which has avoided an effective examination of project cost management difficulties. As project environments are dynamic and constantly evolving, it is crucial to update the project budget regularly to achieve optimal project outcomes.

Earned Value Management (EVM) is a commonly used technique by project managers to track and manage project progress. It involves integrating scope, time, and cost elements of a project to determine how much work has been completed and how much it has cost [15,23]. The EVM can help project managers manage the project's critical elements which are scope, time, and cost more effectively [24]. For many project managers, the use of EVM to forecast project completion time and to

quickly evaluate the expenses of executing scheduled project tasks is an essential tool. Managers can use the anticipated EAC to identify any discrepancies between the planned and actual project expenses in order to fix any fundamental problems. The process of estimating the completion of a project's costs at different stages of its lifecycle is illustrated in **Figure 2.1**.

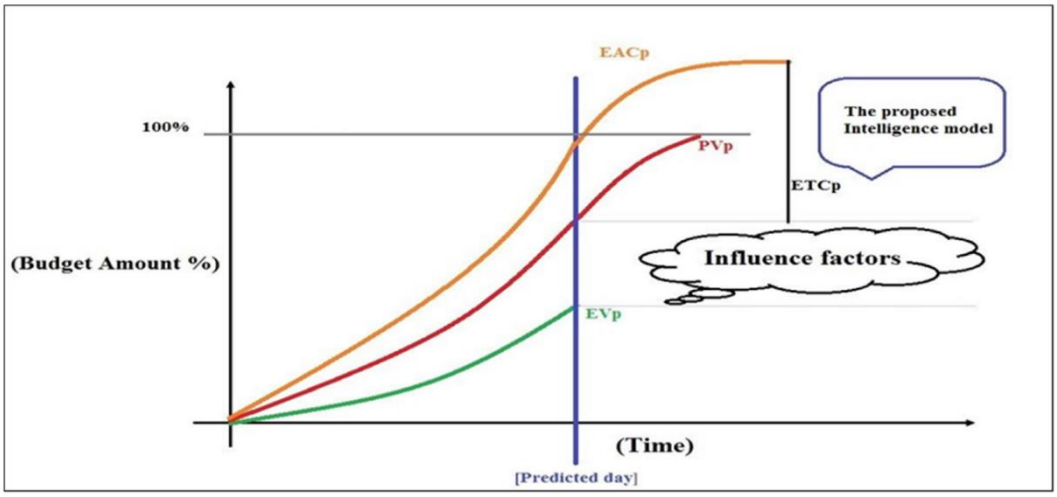


Figure 2.1 The new intelligent model proposed aims to enhance the prediction accuracy of EAC [17].

Practically calculating EAC by formula requires managers to first gather data on project cost management before formulas can be used [25]. One of the biggest drawbacks of utilizing formulas to find out EAC is the large number of different ways you can use it. Therefore, managers have to carefully select the best EAC formula that meets their needs from a variety of options in the literature [26]. The cost of a project is impacted by a variety of factors, and there is a unique characteristic for each project. As a result, the optimum formula for each situation must be chosen. Accordingly, as an alternative to the index-based technique for conducting cost-estimating activities, a range of regression-based approaches have been developed [27,23, 28]. Soft computing models are often utilized to predict the estimation at the end of construction projects. These models typically employ regression analysis to map dependent attribute variables, most commonly the actual project cost, against independent variables, such as time. By leveraging nonlinear modeling techniques, they clarify the relationship between response and predictor variables, offering deeper insights into cost behavior relative to time dynamics.

As building projects are influenced by varying factors and contexts, the development of deterministic models for EAC prediction is often not feasible due

to the high cost involved. In this situation, a cost-effective and rapid approximate inference might be an applicable alternative [29]. In order to extract customized insights from existing data, inference models are utilized, and their methodologies are updated to account for changes in historical data. The ability to deduce new information from past data is a natural function of the human brain. As a result, models capable of simulating the human brain's inference ability could be constructed utilizing artificial intelligence. The core of artificial intelligence suggests that software applications are capable of resolving difficult or poorly organized issues through the use of specialized approaches like fuzzy logic, SVMs or ANN. Given that AI-assisted computer systems can perform human-like tasks, it may be feasible to use AI inference models to solve EAC problems. Whilst attempting to develop artificial intelligence models to address EAC challenges, it has been found that the prediction of EAC is a process that is prone to several uncertainties, one of which is the presence of a significant amount of variable data that is inherent in construction costs. In addition to other elements that affect the cost of a project like productivity at the site, weather, and economical constraints, the researcher's forecast of these things could be time-consuming or nearly impossible. For that reason, predicting models should be able to account for these impacting elements in order to produce an accurate EAC projection [30].

Throughout the construction stage of a project, cost overruns are a common problem Therefore; proactive tracking of project costs is required to detect expected issues. EAC enables project managers in the detection of possible problems and in the planning the most effective solutions to solve them. The most common technique used by project managers in the construction sector to track project conditions and evaluate project performance is EVM [31]. This method's actual mechanism is to set up the exact relationship between projected resources and project objectives. Although EVM is frequently used for project management, it is associated with a number of constraints. Several studies were conducted for improving the fundamental notion of EVM.

Pajares & López-Paredes created two novel ways of identifying project overruns by combining the EVM method and risk management techniques [32]. Regardless of all these efforts for improving the EVM, there are still flaws that necessitate additional work to develop a better solution. The cost monitoring methods were

classified by the researchers into five categories, which include (i) extended EVM analysis, (ii) observational analysis, (iii) statistical analysis, (iv) computerized analysis, and (v) machine learning (ML). Despite being recognized as the most widespread and dependable control system, there are limited number of ML applications available for controlling project performance [25].

Several benefits of early planning to the final results of projects have been demonstrated in the literature [33, 34], but these plans cannot be used universally since they must be evaluated periodically throughout the duration of the project. Thus, efficient project management demands a project plan that is continuously evaluated to reflect the actual project condition. Unless the required steps are taken to ensure proper project management, it will be impossible to repair any problems that are discovered too late. During the first phase of a project, construction businesses routinely prioritize budget planning, ignoring other variables susceptible to change, including cost, information update, and cost management [17].

Managing project cost is crucial in the construction industry, but it is a challenging and time-consuming task due to the impact of several factors on project costs; therefore, it is essential to examine the influence of each of these elements at every project phase [17]. The EAC is still a crucial performance index for the control of cost that requires a precise computation to identify problems and generate suitable actions [35]. The EVM is a widely utilized technique among project managers for calculating EAC [36,37,38]. This technique integrated the scope metrics, schedule, and project cost into a unified measurement structure to assess and measure the actual project status over its baseline, besides, estimating the overall project duration and cost for completion [39]. The EVM is getting more adoption as its capacity to alleviate EVM-related difficulties and enhance its utility increases [40]. Nevertheless, traditional EVM approaches include significant limitations, such as the calculation of the remaining budget using index-based EVM methods is dependent on the use of only old information and performance index [41]. Moreover, early project phases are characterized by limited EVM data, making cost estimates produced by these models unreliable. Even though accurate predictions were able to achieve using typical EVM methods in certain specialized projects, the majority of project cases contained obvious flaws, leading to scenarios in which

firms may not be informed of the best appropriate strategy to use for their project prediction.

The EVM is additionally constrained by the requirement to manually execute reviews due to its application to every project phase. This renders EVM a complex and time-consuming process. Hence, for successful cost control of a project using EVM, it is necessary to automate the engineering management process. Businesses of construction typically utilize powerful computers to examine the first phases of construction projects. However, these computer systems are incapable of responding to changes at every project phase or predicting the EAC of construction projects utilizing the method of EVM. Therefore, the creation of artificial intelligence model capable of resolving issue is highly appealing to project engineering researchers [30].

2.2 Definition of Cost Estimating

Different definitions for cost estimation have been reported, Carr defines an estimate as a declaration of the approximate cost, time, or materials required for executing a construction project [42]. In contrast, Steward argued that the Society for Cost Estimating and Analysis (SCEA) has defined cost estimating as "the skill of assessing the likely cost or value of an activity based on current information." [43]. Furthermore, according to Gao, cost estimation mixes art and science in order to anticipate the subsequent cost of anything based on existing past data that have been changed to account for development new teams, materials, software languages, and technology [44]. With respect to Dysert, cost estimation refers to the method used to predict the cost of the resources and expenses needed for investment alternatives, activities, or projects. Typically, the result of the forecast model, the cost prediction, is utilized to develop the project's budgets " [45]. When it comes to construction contract tendering, Akintoye expressed the view that cost estimation is "a critical aspect of construction contract bidding, providing a basis for projecting the expected cost of resources" [33]. Turochy et al. defined cost estimation as the process of predicting the final cost of a project using data available at a specific point in project development [46]. Hendrickson clarifies that the rough estimate or screening estimate is conducted before the building's design, and thus relies on the cost data of previously built facilities that are comparable to the present

project. A conceptual estimate is usually developed based on the general scope of work and the expected outcome of a project. This estimate is created before the project design is finalized and is based on assumptions and historical data. As the project design progresses, a preliminary estimate is developed based on the completed design drawings and specifications. This estimate is usually more accurate than the conceptual estimate. Finally, the detailed estimate is created based on the final design drawings and specifications, and is used to solicit bids from contractors for the construction of the project [47]. Upon highlights that the Association for the Advancement of Cost Engineering (AACE) International describes cost estimation as the basis for project management, budget preparation, business planning, and cost and schedule control. Included in these expenses are assessments and risk and uncertainty evaluations [48]. According to El Sawy et al., the cost estimate is the firm's lifeblood and is described as the estimation of the amount and the forecasting or predicting, inside one project framework, of the expenses taken to develop and prepare a building [49]. Dysert defines cost estimate as a process of prediction utilized to quantify the cost, pricing, and resources required for a project's scope, investment options, or activities [50].

2.3 Purpose of Cost Estimation

The cost estimate plays a crucial role as a decision-making factor during the initial phase of construction. So, enhanced cost prediction methodologies will allow for improved management of building project time and expenses [51]. In fact, cost estimates are generated and employed for diverse objectives such as assessing the feasibility of a project, competitive bidding, ensuring the proper use of project resources, and so on. The fundamental objective of cost estimation is to generate a dependable and precise estimate of the cost required for project execution. Nonetheless, the projected cost heavily depends on the requirements of the customer and the data and information that are available at the time of estimation [52]. Ostwald states that the objective of a cost estimate is to anticipate the expenditures required to complete a project in accordance with the terms and conditions of the contract's plans and specifications [53]. During the design stage, cost estimates offer additional benefits by enabling the engineer and designer to understand the cost implications of their design decisions. This awareness helps

them to make informed choices that balance the project's cost and quality requirements. Accurate cost predictions can aid managers in making informed decisions about which products are likely to be profitable and which ones may require adjustments [34]. Additionally, accurate cost estimation is crucial during the tendering phase to ensure the sustainability of the business in the long term. It helps to secure a fair price for new contracts, which is essential for profitability and growth. A construction project's bid price must be accurately estimated in order to win the contract and make a profit, despite the fact that the time available for bid estimating may be insufficient in practice. Therefore, it is not always practicable to conduct exhaustive and extensive cost estimates, given that the full cost prediction procedure is both times and cost-consuming. Therefore, when the available time for bid estimating is limited, one of the greatest needs for contractors is a process that doesn’t require more time and could approximate a good bid price [54].

Pajares & López-Paredes described the aim of cost prediction as follows [32]:

- Provides a capital cost evaluation for a selected project component.
- Facilitates the categorization and prioritization of development projects in alignment with the broader business plan. This allows for effective resource allocation and strategic decision-making.
- The cost estimate serves as the basis for project planning and management, defining the expected cost and scope of work.
- By providing crucial information such as hours, resources, tasks, and durations, the cost estimate helps determine the allocation of resources for the project and facilitates the creation of a schedule.
- When resources are more closely aligned with their actual requirements, projects could be easier to control and manage.
- Gives the financial information necessary for constructing a curve of cash flow.
- Could be utilized to evaluate the effect of modifications and to facilitate re-planning.

The main goal of cost estimation is to produce a precise and reliable estimate of the expenses associated with a construction project. This requires a thorough analysis

of all factors that may impact the cost of the project, as well as verification of the accuracy and reliability of the data sources used to inform the estimate.

2.4 Categorization of Construction Expenses

In the report by (Ostwald, Marjuki, and Hinze), the overall cost of a construction project may be broken down into five categories: labor cost, material cost, markup, equipment cost, and overhead costs [53,55,39]. As seen below in **Figure 2.2**.

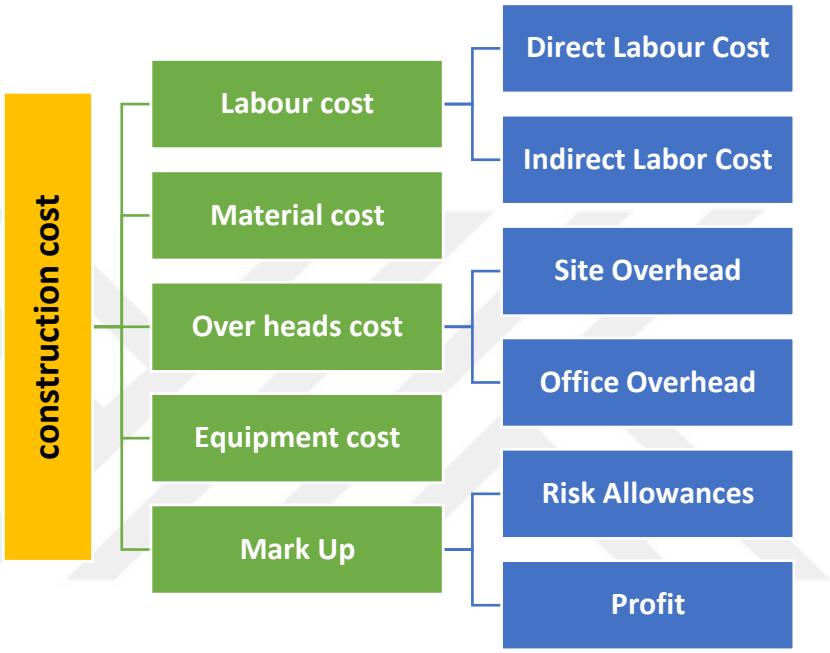


Figure 2.2 Construction Costs Classification

Labor Cost

The portion of construction project expenses associated with labor costs typically ranges from 30 to 50 percent, and can even climb as high as 60 percent of the overall project cost. These costs include both direct and indirect labor expenses that are determined based on their connection to the project.

Material Cost

When estimating the cost of a project, it is crucial to take into account the cost of materials required. This includes not only the direct cost of the material itself but also any additional charges that may be incurred during the procurement process. Such additional costs may include transportation, sales taxes, freight charges, delivery fees, storage fees, and other taxes and losses that may occur during the

handling and transportation of the materials. Taking these factors into account is crucial for accurate and reliable cost estimation.

Overheads Costs

Overheads in construction are costs that are not directly associated with a particular work item. These costs can be substantial, ranging from five to fifteen percent of the project's total cost, depending on how they are included in the cost estimate.

Equipment Costs

The categorization of equipment can be based on its functionality into two types: specialized equipment and general-purpose equipment.

- Specialized equipment is used only for certain building activities and is removed from the work site once its intended purpose is accomplished.
- In contrast, general-purpose equipment is used by all contractors involved in the project and is not limited to any particular work item.

Markup

Markup is an important concept in the construction sector, as it enables firms to allocate an appropriate amount for the recovery of general overhead costs and profit. It involves adding a percentage to the direct cost and estimated overhead cost of a project, which allows the firm to account for additional expenses beyond those directly associated with the job. This markup is essential for ensuring the long-term viability of a business and achieving financial success in the industry.

2.4 Estimator Responsibility

The project failure or success can depend on numerous estimates made throughout the project's timeline, impacting project feasibility and profitability. Estimators play a crucial role in assisting project owners with project planning and budgeting for construction [56]. To make initial cost predictions, the estimator often collaborates with the business department to reach an agreement on the project's scope. The estimator must inform the project team or business department of the expectations for the estimate as soon as possible [53]. It is the estimator's responsibility to make sure that the project team understands the information needs for the cost estimation

process and that the information provided is adequate for producing a reliable cost estimate [45]. The role of the estimator could be to generate an estimate that accurately reflects reality; when the design is completed, they have to possess the foresight to look further than apparent elements and their principal construction expenses [42].

2.5 Estimator Competences

The good contractor on the site can't produce a profit or remain in business if the contract is accepted at a price below the actual cost; therefore, for the best contractor to be a strong competitor in the construction sector, he or she must employ a skilled predictor. According to Carr, Popescu, and Dysert, building an effective estimation process requires employees with significant technical expertise and knowledge. In addition, they describe a set of skills as the following estimation of core competencies [45,57,42]:

- The capability of comprehending and interpreting contract agreements, as well as expertise through having to read plans and specifications for all specifications and specialties.
- Capability to precisely subtract the quantities of building work for which the estimate is being prepared.
- The capacity to see the eventual structure from plans, is a skill that typically requires several years of construction site experience.
- Mathematical, geometric, and statistical knowledge.
- Knowledge of construction techniques.
- In-depth familiarity with estimate software and accessible construction cost database systems.
- Knowledge of team composition, labor output, and the effects of different anticipated site conditions on worker product.
- Capable of working under stress while adhering to all bid specifications and timeframes.
- Possessing office management abilities for managing cost-related project data.

2.6 Types of Estimates for Construction Costs

Estimates are categorized according to the estimation function they serve. However, several factors influence the type of estimate used, such as the purpose of the estimate, the quality and quantity of available information, the level of accuracy required, the calculation methods employed, the time available to produce the estimate, the project phase, and the estimator's perspective [58, 59]. Marjuki reported that the most common types of cost estimates are [55]:

1. Conceptual estimate: A conceptual estimate is an approximation of the costs that is created before the construction drawings are developed. The aim of this stage is to provide a rough estimate of the cost, which is used to determine whether the project is feasible. The accuracy of this type of estimate is usually within a range of -50 percent to +100 percent. It is not intended to provide a precise estimate of the cost, but rather to provide a rough idea of the potential cost of the project.
2. The preliminary estimate is a cost estimate prepared after the completion of the conceptual design and involves defining the scope, schedule, and cost of the project. It is based on established cost data and agreed-upon assumptions and provides the owner with an opportunity to evaluate the design before proceeding to the detailed design phase. The accuracy range of this stage is typically between -30 percent to +50 percent.
3. Engineers estimate: It assists in analyzing bids and depends on a detailed design for which all drawings are complete and created to verify the design is within the available budget. This stage's accuracy ranges from -15 percent to +30 percent.
4. Bid estimate: the contractor's attempt to price the contract during the bidding phase. This stage's accuracy ranges from -5 percent to +15 percent.

As the project progresses from the idea to the final, and bidding stages, a more precise estimate can be created [55]. Cost estimates are classified as conceptual, semi-detailed, or detailed by some researchers, with error rates ranging from 20% for conceptual estimates to 5% for detailed estimates [53].

In accordance with the AACE Recommended Practice and Standard (AACE, 1990), three types of estimates are identified based on the level of project specification [60]. They are:

1. Rough Order of Magnitude estimate with an accuracy range of -50% to +100%
2. Preliminary estimate with an accuracy range of -30% to +50%
3. With an accuracy range of (-5% to +15%), this is a definitive estimate.

Broadly speaking, construction projects have two categories of estimates: rough estimates (also referred to as conceptual, preliminary, or budget estimates) and detailed estimates (also referred to as final, definitive, or contractor's estimates). Conceptual estimates are generally less accurate, with a margin of error between -15 to +30 percent, while definitive estimates are highly specific and have an accurate limit of -5 to +15 percent [56].

2.7 Indexed Formulation for EAC

To create the EAC, the data collected from the contractor's cost control reports, such as the report of cost performance or the report of schedule/cost condition, can be utilized [61]. The reliability of this data is determined by the extent to which the contractor adheres to a system of internal control that includes budgeting, scheduling, and contractual effort analysis. The calculation of EAC depends on various data parameters collected from the report of cost control, which include the budgeted cost of work scheduled (BCWS), ACWP, and budgeted cost of work performed (BCWP). These data parameters are usually reported on a monthly basis. The cumulative/average values are then calculated for the duration of the contract [26]. The estimation formulas for EAC are classified as the index, regression methods, and artificial intelligence modeling. The index-based formula is denoted by the equation (2.1).

$$EAC = ACWPc + (BAC - BCWPc)/Index \quad (2.1)$$

Where "c" denotes the data accumulation. The BAC represents the overall project budget that has been approved. The index technique, which frequently incorporation of BCWS, BCWP, and ACWP, is utilized primarily for detecting the cost of the residual project work (BAC - BCWPc).

The adjustment is predicated mostly on the premise that the project's previous schedule and cost performance is predictive of its future performance. performance indices have Four categories:

$$CPI = BCWP/ACWP \quad (2.2)$$

$$SPI = BCWP/BCWS \quad (2.3)$$

Schedule Cost Index (SCI): $SCI = SPI * CPI$ (2.4)

Composite Index (CI): $CI = W_1 * SPI + W_2 * CPI$ (2.5)

W_1 and W_2 which are weights in equation (2.5), could be any value between 0 - 1 and often add to 1.

Using data that can be measured monthly, on average, or cumulatively, various indices can be calculated. The following conventions can be employed for labeling purposes: "CPIx" = an average index over x months; "CPIc" = a cumulative index; and "CPIm" = an index based on the most recent month, with calculations starting from the most recent month and extending to the furthest month. For instance, CPI_3 is defined as a three-month CPI average derived from the present month and the two most recent months. Similar notation was adopted for the SPI, where the same labeling convention was used. For instance, "SPI6" represents the average SPI over a six-month period starting from the current month and ending with the previous five months. There are two methods for averaging these indices; A ratio of sums over x months is typically the averaged index [62]:

$$CPIx = \sum BCWPx/ACWPx \quad (2.6)$$

$$SPIx = \sum BCWPx/BCWSx \quad (2.7)$$

Alternatively, the total monthly indices can be divided by the number of months to calculate it [34]:

$$CPIx = \sum CPIm/\ x \quad (2.8)$$

$$SPIx = \sum SPIm/\ x \quad (2.9)$$

The second and third classes of EAC formulas are referred to as the "regression" and "other" formulas, respectively. The regression-based formulas are derived using linear or nonlinear functions. This study employed nonlinear regression to analyze the nonlinear relationship, even if it can be transformed into a linear relationship. Typically, the independent variable (BCWP) and the dependent variable (ACWP) are the performance time in all other scenarios. Clearly, the EAC formula contains potentially unlimited numbers. The expert has discretion over which set of formulas to apply. In 1991, a performance analyzer was utilized, which

is a well-known software program that allows the analyzer to select from many formulas. However, there is no instruction regarding the precise form of a formula [30].

2.8 EAC Computation Applying the Empirical Formula

The empirical formulation method for EAC prediction is a systematic method that utilizes historical data and takes into consideration the current project performance to estimate the final cost of a project. This type of formulas presents a novel modeling approach that takes a step away from the use of statistical or advanced heuristic methods, nonetheless, it cannot be seen solely as comparative analysis [63]. In this method, by examining trends and patterns in project execution, including factors like cost overruns or schedule delays, empirical formulas are tailored to their specific project environment. Empirical studies generally proposed a CI which composes different PIs, especially CPI and SPI with weights, to calculate EAC [64, 65]. The development of these techniques is generally based on the subjective assignment of index weights. However, these index weights can change throughout the project, therefore different approaches are proposed to determine these weights. For instance, Lollar [64] introduced a method for determining the weights associated with the cumulative SPI and CPI. These weights are calculated based on the proportion of percentage changes in cost and schedule values relative to their total variations. Similarly, Parker [65] proposed a different approach that involves the calculation of a sequence of composite indices through the gradual adjustment of weights ranging from 0 to 1 in increments of 0.1. The project's specific conditions determine the choice of the most fitting CI. Exemplified through a numerical illustration, a simulation-based method that subjectively evalutes different EACs is advocated for by Holeman [66]. As a result, a special approach – that deviates from conventional EVM index-based approaches - has been introduced for CEAC. Totaro [67] suggested a technique for establishing weights for composite indices grounded in the percentage completion offering a departure from conventional approaches. The analyst initiates the process by determining the initial weights of the SPI and CPI, considering project-specific factors such as predetermined manpower loading. Batselier and Vanhoucke [68] proposed a model, which was developed using the exponential smoothing

forecasting (ESF) approach, in which they introduced a new performance factor that is based on a smoothing parameter, β. This parameter, which ranged from 0 to 1, allows for adjustments based on recent progress or corrective actions by dictating the responsiveness of forecasts to current project performance. Thus, their model offers both static and dynamic approaches. Khamooshi and Abdi [69] also used ESF to calculate the earned duration index which is one part of their technique, called earned duration management (EDM). In this technique, new performance metrics have been introduced to develop new formulas, providing a comprehensive view of schedule attainment [70]. Furthermore, Lipke, et al. [71] proposed a study that integrates elements of EVM and Earned Schedule (ES) with statistical prediction and testing methodologies. Specifically, it entails the statistical analysis of CPI derived from EVM and SPI(t) from ES.

There is a vast array of EAC formulas that can be generated because there is an infinite number of weight combinations that can be used in the computation of the CI. Consequently, selecting the right formula or formulas may pose a challenge for the analyst. The user gave permission to select from a range of calculations by the popular analytical program performance analyzer [72]. Nevertheless, there is no recommendation given as to which formula or collection of formulas is the most correct. Covach et al. analyzed data from 17 Navy contracts, including 14 development contracts and 3 production contracts, to investigate twelve different formulas based on indices. The indexes were generated using single, average, and cumulative data [73]. **Table 2.1** presents a summary of the results obtained from comparing different index-based formulas. The choice of the most accurate formula varied depending on the project's level of completion. The CPI3, which is the average CPI for three months, was determined to be precisely independent of the state of completion [73].

Table 2.1 EAC Comparisons results (13 Production and 14 Development Contracts (Navy))

Best Performing Formula/Model	**Completion Stage**
CPI3, CPIc, SCIc	(0-40%) Early
CPI3, CPI6, CPIc, SCIc	(20-80%) Middle
CPI3, CPI6, CPI12	(60-100%) Late

Bright and Howard analyzed data from 11 development contracts managed by the Army and investigated 11 different formulas and models [74]. There are nine formulas based on indices (CPIc, SPIc, SCIc, CPI3, CPI6, CPI12, SPIcxCPI6, 0.75CPIc + 0.25SPIc, 0.5CPIc + 0.5SPIc). **Table 2.2** summarized the results. Bright and Howard discovered in the initial stages of their study that the composite indices were the most precise among the different formulas and models they analyzed. In the early phases of the investigated contracts, composite formulas with greater weights for SPI were more accurate, indicating a decline in the SPI's information content. The average CPIs were most precise throughout the middle phases. When the middle stages of a contract see a significant increase in CV, Bright and Howard hypothesized that an index averaging over a smaller time frame could be more reliable than one averaging through a longer time frame. In later stages, CPIc and SCI accuracy improved. Furthermore, during the initial phases of the analyzed contracts, it was discovered that the SCI served as a reasonably precise index [74].

Table 2.2 Comparison of EAC Results from 11 Development Contracts Administered by the Army

Top-Performing Formula/Model	**Completion Stage**
Composite, SPIc, SCI	(0-40%) Early
CPI3, CPI6, CPI12	(20-80%) Middle
CPIc, SCI	(60-100%) Late

Riedel & Chance examined 6 index-based models utilizing information from 56 contracts (40 productions, 16 development) maintained by the Air Force [75]. Six different formulas, namely CPI3, CPIm, SCIc, CPIc, 0.2SPIc +0.8CPIc, and (X)CPIc + (1-X)CPIc (where X represents the percentage of completion), were evaluated at four different completion phases. Additionally, the sensitivity of the results to the type of weapon system was also analyzed, with a focus on five engines, five avionics, and eight aircraft. In most cases, Production contract EACs become more precise than development contract EACs. **Table 2.3** presents a concise overview of the detailed results. The abbreviation PC denotes the formula that dynamically adjusts the weights of the CI according to the percentage of work completed. The notation "20/80" signifies an 80% weight assigned to the CPIc component and a 20% weight assigned to the SPIc component within the CI [75].

Table 2.3 EAC Comparisons Results (40 Production Contracts and 16 Development (Air Force)

Phase	System	Completion Stage				Overall
		25%	50%	75%	100%	
Development	Aircraft	SCIc	CPI3	CPI3	20/80	SCIc
Production	Aircraft	SCIc	CPI3	SCIc	CPIc	SCIc
Development	Avionics	SCIc	CPI3	SCIc	CPIc	CPI3
Production	Avionics	20/80	SCIc	20/80	SCIc	20/80
Development	Engine	CPIm	SCIc	CPI3	CPI3	CPI3
Production	Engine	PC	CPIc	SCIc	PC	CPIc

2.9 Regression-Based Model for EAC Computation

The regression-based model serves as a statistical technique employed in data analysis to establish connections between a dependent variable and one or multiple independent variables. The main purpose of using regression analysis is to forecast the overall project cost (dependent variable) —especially in the context of construction project management and EAC. This takes into account the diverse input factors or variables (independent variables).

In a study written by Skitmore and Ng [76], the researches utilized a forward cross-validation regression analysis approach to forecast the predict expenses and durations in construction projects. The results of their research suggest that by several variables (like the total contract amount) significantly influence the cost forecasting accuracy. They argue that their model has a lot of potential when it comes to forecasting the cost of construction projects (and thereby helping the managers mitigating cost-related risks). Another researcher, Swei, et al. [77], combined Maximum Likelihood (ML) and Least Angle Regression (LAR) in order to predict road pavement costs and suggested that their model could be further improved by putting in more cost-related variables. They also advised for the use of AC data (instead of bid prices) to further enhance their model. Furthermore, Li, et al. [78] examined regression analysis as a means of predicting construction project expenses. Through a stepwise variable selection method, the authors thoroughly analyzed the relationship between seven independent variables and

project costs, resulting in improved model accuracy compared to traditional methods. Also, Watkins [79] presented a new approach that integrates linear regression analysis and an adaptive iteration of the Rayleigh Norden model to forecast the EAC. He argued that the Rayleigh-Norden model ably captures the life cycle dynamics of workforce changes observed in defense contracts. The methodology entails linear regression analysis on the ACWP plotted against time, using quarterly data from three contracts outlined in Cost/Schedule Status Reports. Additionally, Narbaev and De Marco [28] introduced a novel approach to determining the cost of uncompleted project phases. Their approach advocates for employing a modified Gompertz growth model and nonlinear regression curve to analyze data pertaining to each project. Furthermore, the researchers incorporated the earned schedule metric to serve as a cost performance indicator. In their equation, a factor representing the projected project completion time has been integrated. Therefore, when compared to the four traditional indexed-based formulas, the method they used has proved to be more effective when it comes to generating cost estimates throughout all the stages of the project —from early stages to late stages. Furthermore, a novel nonlinear CEAC methodology that employs regression analysis was proposed by Narbaev and De Marco [23]. Their approach incorporates ES principles (in addition to four distinct growth models (Bass, Weibull, Gompertz, and Logistic)). Their study found that their model outperformed the other three models and the index-based approach, after thoroughly analyzing CEAC error values and conducting statistical validity assessments. This was particularly evident during the early and middle stages of a project's completion. Furthermore, Ottaviani and Marco [80] conducted another study using multiple linear regression (MLR) analysis to forecast EAC. Although their methodology is simpler compared to other models developed using regression analysis, they still claim that their model's performance is superior to the standard formulas in terms of accuracy and variance. Additionally, Ayalew et al. [81] address the inability of construction industries to complete public building projects on time and within the budget. The study tries to assess certain ways in which the regression modeling might help improve the accuracy of the EV analysis in the prediction of project performances. The research collected raw data from ten such building projects between the years 2019 and 2022 and resorted to statistical tools in developing a regression model. This resulted in very accurate predictions of the cost

and schedule performance indexes, as expressed by the correlation coefficient, which was between 88.80% and 96.60%, while the coefficients of determination lay between 78.90% and 93.20%. The study provides the conclusion that regression analysis is a practical method for making predictions regarding EV indexes in construction projects. Furthermore, Chacko has presented a time-series prediction technique called "Adaptive Forecasting" that requires five months of data to generate an accurate EAC estimate [82]. As each new monthly data point is obtained, the adaptive forecasting algorithm evolves accordingly, allowing it to provide short-term EAC predictions. On the other hand, Watkins suggested a technique for EAC prediction by utilizing linear regression analysis along with an adaptive version of the Rayleigh Norden model. Furthermore, El-Sabban proposed a new approach to estimate the EAC using Bayesian probability theory [83]. The method starts by assuming a normal probability distribution, a mean, and a standard deviation at the beginning of a project for EAC.

2.10 Soft Computing Models for EAC

Typically, professionals rely on their intuition to create cost estimates for building projects. However, to improve the accuracy of these estimates, it is crucial to develop and apply scientific methods during project planning and design. It has been proposed that the accuracy of cost estimations can be enhanced using a novel evolving fuzzy neural network [84]. A novel intelligent prediction technique has been developed by incorporating the strengths of neural networks, genetic algorithm (GA)s, and fuzzy logic into a unified approach. The intelligent forecasting system developed in this study is a powerful tool for solving complex problems and identifying optimal solutions. The researchers concluded their investigation by introducing two types of predictors that can be used to generate precise cost forecasts for conceptual building projects in their early stages.

Given the complexity and uncertainty of construction projects, construction decision-makers must operate in a challenging environment. The process of making decisions based on limited information and intuition is inherently flawed, as it relies heavily on the experience of the decision maker and the quality of available information. However, there are many factors that can affect the accuracy and reliability of this information. To address this issue, an alternative approach has

been developed in the form of the evolutionary fuzzy support vector machine inference model. This model aims to more effectively utilize and retain acquired knowledge in order to make more accurate and informed decisions [85]. The study identified two simulated and actual problems of construction management and showed that the hybrid intelligent method provided is a viable means of resolving many construction-related issues.

Throughout project feasibility analyses, estimations of conceptual cost and their effect on the success of the final project are critical. This is because these approximations are essential in providing useful insights that can assist in evaluating the project, formulating engineering blueprints, organizing, and managing project costs the improvement of conceptual cost estimation precision was proposed by [22] through the use of a hybrid evolutionary fuzzy hybrid neural network (EFHNN) model. To achieve this, they combined neural networks (NN) and high-order neural networks to form a hybrid neural network (HNN). The hybrid network developed by [22] uses a unique approach where linear and nonlinear neuron layer connections alternate to increase the accuracy of conceptual cost estimation. Additionally, they incorporate fuzzy logic (FL) into the HNN to manage uncertainties, resulting in a fuzzy hybrid neural network (FHNN) that can provide more precise cost estimates. The Evolutionary Fuzzy Hybrid Neural Network (EFHNN) employed in this study is appropriately named as such because it incorporates a GA to optimize the hybrid FHNN, which is used in both the HNN and FL. The main aim of this study was to analyze and compare the overall and categorical expenditures of completed projects. The EFHNN, which was proposed and implemented in the study, can be used as a reliable cost estimator during the initial phase of project planning. Interestingly, it was observed that the hybrid network architecture with nonlinear and linear neuron layer connectors used in EFHNN outperformed single linear NN models.

EAC is an important metric used by project managers to predict the final cost of a project, enabling them to assess the project's risk and performance. Though high-level project decisions are typically made by executives, their limited technical expertise can lead to inaccurate decision-making. In [19], a novel approach for estimating EAC is proposed, which involves using a combination of Gaussian process (GPr) and particle swarm optimization algorithm (PSO) in an evolutionary

predictive model. The proposed model also incorporates an AI method to simulate human decision-making in addressing management challenges. The proposed hybrid model in this work utilizes the GPr to determine the input-output relationship and the PSO algorithm to optimize the hyper parameters in the data function. Its viability was examined for determining EAC and calculating the probability of a change in the forecasted model monitor. The model imitates the human decision-making process and offers a promising solution to solve management difficulties. The study shows that the proposed model can estimate EAC effectively, which could greatly benefit project managers aiming to improve their project cost control efficiency. The accuracy of the model was verified using real-world data from construction projects.

Feylizadeh et al. have proposed a novel approach to EAC prediction, incorporating both qualitative and quantitative elements using a fuzzy neural network model. Unlike conventional methods that rely on prior performance factors, this model provides a more accurate estimate of current project status and future event costs. The effectiveness of this approach was tested in a real-world case study, yielding promising results. [86]. On the other hand, using a Bayesian method according to expert opinion can allow the exploitation of subjective judgments officially and rigorously; this can result in a more accurate estimate of completion within the EVM framework. Caron et al. proposed a Bayesian network model that combines project data and expert knowledge to create a predictive and supportive tool for process improvement. The model improves the project management decision-making process and is applicable in all project phases, especially in the early stages when data is incomplete. It provides confidence intervals that reflect potential future project situations. The efficacy and applicability of this model were demonstrated and certified in the gas and oil industry [87]. The research outcomes also suggest that the Bayesian model can be adapted to improve project management processes in various other industries. To enhance the ease of use and input-output management process, the Bayesian model has been transformed into a software package. On the other hand, Enshassi, Mohamed, and Abushaban conducted a study in the Gaza strip to evaluate the factors affecting contractor performance in building projects. The study utilized a questionnaire to gather input from project participants, including owners, consultants, and contractors. The

results indicated that poor project performance in Gaza was mainly attributed to factors such as inadequate equipment, lack of experience, poor quality materials, change in material prices, lack of leadership skills, lack of resources, and the closure of borders [88]. Furthermore, Wiguna & Scott's research on Indonesian construction projects, the main factors contributing to cost and time overruns were identified as inflation, design revisions, poor design, weather conditions, and payment delays. They conducted an analysis of 22 building construction projects, and the contractors involved in those projects were surveyed to obtain their opinions on the most significant variables affecting project performance [89]. In addition, Mahamid conducted a questionnaire study with eighty-four contractors in the West Bank of Palestine to investigate the primary causes of contractor business failure. The study identified forty-four factors that were categorized into three groups: financial, managerial, and external. Mahamid found that the most influential factors were the increase in construction material prices, payment delays, lack of experience, slim profit margins, and the political climate. [90]. Additionally, according to the research conducted by Frimpong and colleagues, the leading factors that affect the timely completion and cost performance of construction projects in Ghana are contractor financial difficulties, inadequate management, and substandard technical performance [13]. Furthermore, Luu & Kim calculated the overall construction costs for projects in Vietnam using the ANN. The input variable was defined by sending out 91 questionnaires, of which only 14 were returned. The acquired data were processed and prepared for neural network training and testing. Using MATLAB software, the ANN was designed. To apply the neural network to practical problems, Visual C++ was used to construct the program. The findings demonstrate that it is possible to anticipate the overall building costs for residential developments in Vietnam and that neural networks are effective cost models [91].

Xiao-Chen developed a backpropagation-based ANN model to estimate building costs of highway projects. The study utilized data from road projects completed in certain regions between 2000 and 2002. The data was separated into seventeen standard training projects and sixteen standard testing projects. The input variables for the model consists of nine variables that influence the development of highway projects: highway grade, type of foundation treatment, width, height, cross-section

type, half-filling, cutting, half-digging, and embankment. On the basis of these variables, the gathered projects were categorized, and the qualitative data were converted into quantitative data for use in the model. The test results indicate that there is a small margin of error between the actual and predicted values, with a relative error of less than 5 percent demonstrated by the model. It is feasible to meet the need for cost estimation and predict the installation cost of highway projects. It demonstrates the model's ability for generalization and precision. In addition, it is suggested that it is totally conceivable to implement ANN technology in highway projects in order to predict accurate outcomes [92]. Furthermore, Roxas & Ongpeng created an ANN model to forecast the total cost of building projects in the Philippines, utilizing MATLAB software. A feed-forward backpropagation algorithm was utilized to train and select the model. The study utilized a dataset that consisted of 30 projects, with 60% allocated for training, 20% for validation, and 20% for testing purposes. The model incorporated six input variables, namely reinforcing steel weight, concrete volume, number of floors, concrete volume, number of basements, and formwork area which were deemed to be the most influential aspects or characteristics influencing the building cost in the Philippines. As reported in [93], an effective model structure was chosen by selecting six input variables, utilizing a hidden layer that had seven neurons, and having a single output node that represented the construction cost. The model's ability to accurately estimate building structure costs was validated through training and testing. Moreover, a research study by Cheng et al. presented an evolutionary EAC model that outperforms current commonly used formulas in generating dependable cost estimates for projects. An evolutionary support vector machine inference model (ESIM) was produced using a learning model that combined fast messy genetic algorithm (fmGA) and SVM. The ESIM was utilized to estimate the final project expenses by applying it to previous cases. The investigation of the gap between estimated and actual project costs using EAC estimate, project cost diagrams, and project cost influence indices was conducted to identify potential issues and support project managers in controlling project costs more efficiently. The learning outcomes were validated through real-world applications, which demonstrated the effectiveness of the training models. The provision of reliable EAC trend predictions to project managers assists them in efficiently controlling project costs and taking preventive measures to address potential issues [17]. Also, Narbaev and

Marco developed an innovative CEAC approach that utilizes a revised index-based equation to anticipate the estimated expenses for the remaining work by fitting a nonlinear regression curve to the Gompertz growth model. In addition, the proposed equation includes scheduling progress as a cost-performance factor. In order to accomplish this, it incorporates a factor based on the Earned Schedule that indicates the expected duration to completion. The proposed methodology by Narbaev & Marco has been shown to outperform four traditional index-based formulas in early, middle, and late-stage estimates. The model offers a practical solution for Project Managers to accurately calculate CEAC by integrating progress status, and the research is a valuable addition to the field of EVM, capturing the complex relationship between schedule and cost factors [28]. On the other hand, the study conducted by Cheng et al. presented a novel approach to improve EAC predictions, utilizing a combination of different techniques. The proposed model utilizes fuzzy logic to address uncertainty and improve reasoning capacity, while a weighted support vector machine (WSVM) is used for time series data analysis. Furthermore, a fast messy genetic algorithm (fmGA-II) is used to optimize model tuning parameters. The simulation results showed that the proposed model improved the accuracy of EAC forecasting considerably [35]. On the other hand, Cheng & Roy suggested a research study that aimed to design an evolutionary fuzzy support vector machine inference model (EFSIM) to combine fuzzy logic (FL), SVM, and fast messy genetic algorithm (fmGA). This AI hybrid system offers an innovative method for maintaining and using practical knowledge. The effectiveness of EFSIM was verified through two simulations that addressed actual construction management problems, indicating that EFSIM is an efficient tool for solving various issues in the construction industry [85]. In addition, Feylizadeh et al., introduced an innovative approach that accounts for qualitative and quantitative factors that impact EAC prediction. They developed a fuzzy neural network model that accurately forecasts all future durations in addition to the project duration. The efficacy of this model was tested on a real-world case study, and preliminary results showed promising outcomes [86]. On the other hand, AlHares & Budayan developed innovative coupled intelligence models, integrating global harmony search (GHS) and brute force (BF) with extreme learning machine (ELM) techniques to enhance the accuracy and efficiency of EAC estimation. GHS and BF identified influential attributes for the EAC variable. A traditional ANN served as

a benchmark for validation. Historical data from UAE construction projects were used to test the models, focusing on determining EAC and assessing forecast model changes. The goal was to provide reliable EAC estimates, aiding project managers in improving cost control. Results showed significant improvement with GHS-ELM and BF-ELM over traditional and hybridized ANN models [10].Furthermore, Kamoona & Budayan introduced a new intelligent model, the deep neural network (DNN), for EAC calculation, validated against the SVR model. The study examined fifteen projects in the Iraq region using historical project information to showcase the model's capability in engineering applications. The second phase integrated input optimization algorithms, GA and brute force algorithm (BF), to approximate input attributes and explore highly influential factors in EAC calculation are used. This study was designed to develop an intelligent model for precise project cost estimation, surpassing traditional methodologies. Additionally, it sought to introduce an effective methodology for the control of project costs. The proposed GA-DNN has been demonstrated to be a reliable and robust intelligence model for the calculation of EAC [94].

3
METHODOLOGY

The main goal of this chapter is to provide a general view of a hybrid intelligent model that incorporates the FFA and AOA to ANFIS and ANN for prediction of EAC. The study also compares the hybrid models results with the traditional ANFIS and ANN models, which are data-driven machine learning approaches.

3.1 Overview of Adaptive Neuro-Fuzzy Inference System (ANFIS)

In recent times, the combination of fuzzy logic and neural networks has proven to be a successful soft computing technique. Among these techniques, the ANFIS is a popular model that has shown to provide more robust results than other fuzzy inference systems due to its flexibility, adaptability, and efficiency in handling complex issues with highly non-linear data [95]. ANFIS is a widely used soft computing technique that combines neural networks with fuzzy logic. A critical aspect of ANFIS is the selection of appropriate membership functions and shapes, which require expert knowledge of fuzzy logic. The computational cost of an ANFIS model is also affected by this selection. The number of rules in the ANFIS rule base is determined by the number of membership functions for input space segmentation for each input, which further impacts the effectiveness and computational efficiency of the ANFIS-based model [96]. Researchers have the flexibility to choose from a broad range of membership function shapes for the ANFIS-based model. These shapes can be Gaussian, triangular, bell, or a combination of them, depending on the degree of fuzziness in the input data. Such customization is necessary to improve the accuracy of the model and make it more precise [97].

3.1.1 Membership Functions

Membership functions are a key element of fuzzy set theory, allowing for the representation of partial truth values between 0 and 1. They are particularly useful in addressing real-world problems with ambiguous or imprecise parameters. Each membership function is associated with a curve that maps the input space to the universe of discourse, and is given a specific name based on its shape, such as triangular, trapezoidal, bell-shaped, or Gaussian. **Figure 3.1** displays the basic forms of membership functions.

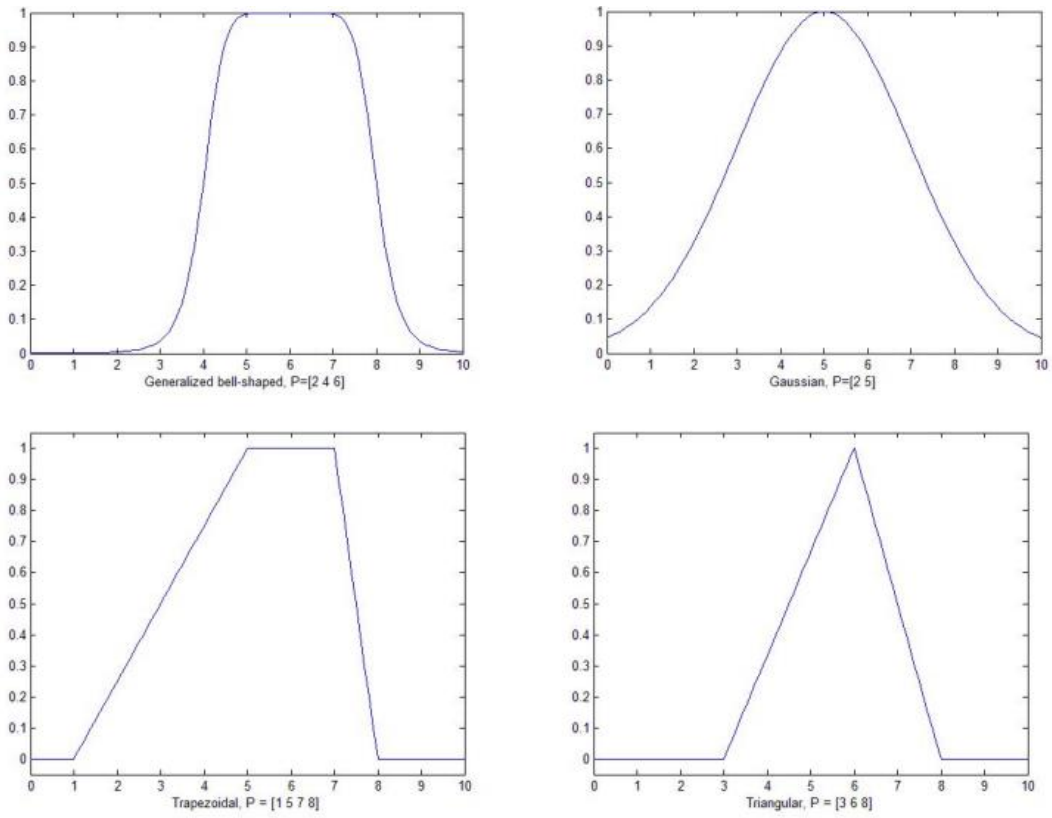


Figure 3.1 Membership functions basic shapes [98].

The Trapmf, a trapezoidal membership function, is a curve defined by four scalar parameters: a and b for the feet and c and d for the shoulders. Its mathematical expression is represented by equation (3.1).

$$\text{trapzoidal}\,(x,a,b,c,d) = max\left(min\left(\frac{x-a}{b-a},1,\frac{d-x}{d-c}\right),0\right) \qquad (3.1)$$

The membership function known as Gbellmf has a symmetrical bell-like shape. This function is defined by three parameters, as shown in Equation (3.2): a

determines the curve's width, b is a positive integer, and c determines the curve's center in the input space.

$$\text{bell}\,(x, a, b, c) = \frac{1}{1 + \left|\frac{x - c}{a}\right|^{2b}} \tag{3.2}$$

The Trimf, which is the simplest form of membership function, has a triangular curve. It requires three parameters that are defined by three points on the curve: a and c represent the feet, while b represents the peak of the curve. The function's mathematical formula is expressed by Equation (3.3).

$$\text{traingular}\,(x, a, b, c) = max\left(min\left(\frac{x - a}{b - a}, 1, \frac{c - x}{c - b}\right), 0\right) \tag{3.3}$$

Gaussian (Gausmf) also has a smooth curve, like the bell-shaped membership function. However, among the aforementioned three membership functions, it employs only two parameters: c for locating the center and for defining the breadth of the curve as mathematically stated by Equation (3.4). [98].

$$\text{gaussian}\,(x, c, \sigma) = e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2} \tag{3.4}$$

3.1.2 Construction of ANFIS

Adaptive network is a type of ANN architecture, in which the neurons are organized in a feed-forward network with directed links connecting them. Unlike conventional neural networks, adaptive networks do not assign fixed weights to the connections between nodes. Instead, they allow the network to adapt its connections and adjust its structure to optimize its performance on a given task. Each node in the network receives input signals, performs a computation on those signals, and produces an output, which is passed on to other nodes in the network. In an adaptive network, the nodes are interconnected by directed links, and each node processes the input signals to produce an output. The node functions are usually parameterized, and their parameters can be adjusted to modify the overall behavior of the network. By changing these parameters, the adaptive network can be modified to suit the specific requirements of a particular task. The system architecture, illustrated in **Figure 3.2**, consists of five layers: the fuzzy layer, product layer, normalized layer, de-fuzzy layer, and total output layer. The ANFIS approach creates a fuzzy inference system (FIS) model using input/output data for a specific set of parameters, and tunes the

membership function parameters using either a backpropagation algorithm or a combination of a least squares-type method and backpropagation. The ANFIS is designed to find the most suitable values for the equivalent fuzzy inference system parameters through the use of a learning algorithm. During the training phase, the optimization of parameters aims to minimize the gap between the expected and actual output. The ANFIS premise parameters are the ones that need to be optimized, and they determine the shape of the membership functions [99]. Various optimization techniques can be utilized to minimize the error measure after designing the membership functions. The parameterization of an adaptive network facilitates the learning process of fuzzy systems from the input data. The rule base of a two-input (x and y) and one-output (f) ANFIS system is expressed in Equations (3.5) and (3.6): consequence premise parameters

$$\text{Rule 1: If (v is } V_1\text{) and (d is } D_1\text{) then } f_1 = p_1 v + q_1 d + r \quad (3.5)$$

$$\text{Rule 2: If (v is } V_2\text{) and (d is } D_2\text{) then } f_2 = p_2 v + q_2 d + r_2 \quad (3.6)$$

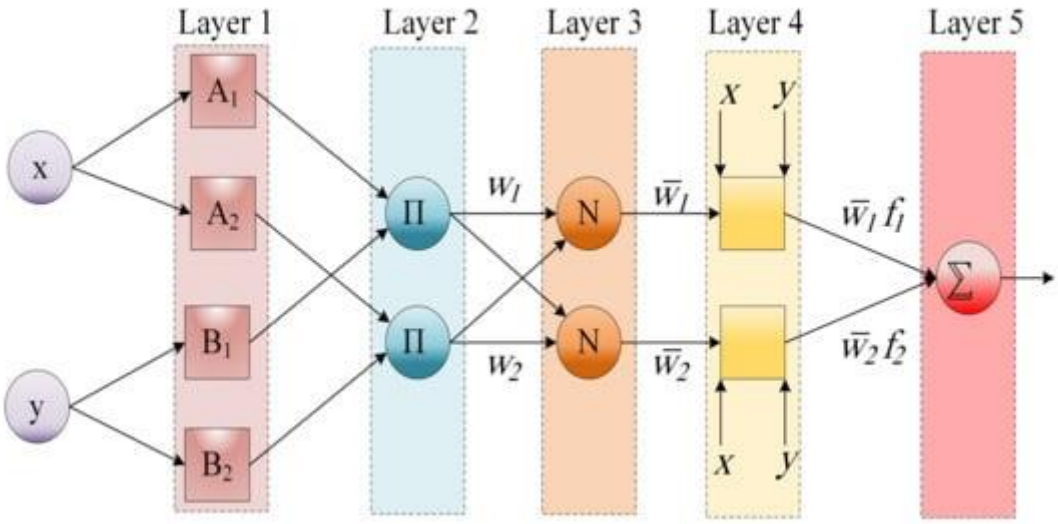


Figure 3.2 ANFIS framework[101].

The ANFIS (Adaptive Neuro-Fuzzy Inference System) comprises both linear and nonlinear parameters, namely p_1, p_2, q_1, q_2, r_1, r_2, V_1, V_2, D_1, and D_2. While p_1, p_2, q_1, q_2, r_1, and r_2 are the linear parameters of ANFIS (consequent), V_1, D_1, V_2, and D_2 are its nonlinear parameters that signify membership functions. The consequent parameters namely p_1, q_1, and r_1 [100]. Both a circle and a square are employed to represent adaptability. In ANFIS, a circle represents a fixed node, whereas a square represents an adaptive node. The ANFIS model combines fuzzy logic and neural

networks to adapt its parameters during training. It is important to carefully select the number of training epochs, membership functions, and fuzzy rules to avoid overfitting or underfitting the data. A hybrid approach that combines the least-squares method, gradient descent, and mean square error (MSE) can be used to adjust these parameters. Minimizing the training error during the training process is crucial for achieving a precise ANFIS system output. A smaller difference between the ANFIS output and the desired objective indicates a more accurate ANFIS system [101]. The fusion of fuzzy logic and neural network has given rise to the concept of fuzzy neural network (FNN), which has been developed and utilized. ANFIS is a term that refers to the combination of fuzzy logic and neural network, which is the organization of this approach. While the neural network has multiple inputs and outputs, fuzzy logic has multiple inputs and only one output, which is why their integration is referred to as ANFIS.

The ANFIS model used in this study is a hybrid computational framework that integrates the flexibility of neural networks and the interpretability of fuzzy logic. A comprehensive description of the ANFIS model used is described below:

1) Layered Architecture: a layered architecture is employed within the ANFIS system, which consists of input, fuzzy inference, and output layers. The systematic processing of input data is facilitated by the proposed architecture which employs fuzzy logic inference mechanisms. This in turn produces accurate predictions. Each layer contributes to the overall functionality and predictive power of the model by performing context-specific computations.
2) Fuzzy Inference System: fuzzy inference systems are utilized in the ANFIS system to model linguistic variables and perform fuzzy logic processes. Furthermore, they utilize Gaussian membership functions to “fuzzify” the input data and generate fuzzy inference rules.
3) Rule-Based Rule Activation: ANFIS employs rule-based activation to identify the effect of each fuzzy rule on the final output. The activation function calculates the degree of activation for each rule based on the input data and fuzzy membership functions. This enables the model to quantify the contribution of individual rules in the inference process.
4) Parameter Learning: The model uses gradient-based optimization algorithms - such as backpropagation and the Adam optimizer – to learn the parameters

associated with fuzzy inference systems. Moreover, the model adjusts the means and standard deviations of the Gaussian membership functions to minimize the loss function.

5) Training Process: To reduce the disparity between predicted and actual outputs, the ANFIS model undergoes a training procedure through which it gradually modifies its parameters systematically (through iterations). This training process incorporates both forward and backward passes within the network, with the computed gradients determining the parameter updates.
6) Evaluation: Throughout the training process, the model evaluates its performance by utilizing the use of several indicators —like Mean Absolute Error (MAE), MSE, and R-squared value. The evaluation involves a comparison between the predicted and actual outputs allowing it to measure the predictive accuracy of the model.
7) Loss Function: The Huber loss function is used to calculate the discrepancy between the forecasted output of the ANFIS model and the genuine target values in the training phase. The function's role is to guide the optimization algorithm (specifically the Adam optimizer) towards efficiently minimizing the difference between predicted and actual values. The Huber loss function is a robust regression technique and it is less influenced by outliers, unlike the MSE loss function, which is susceptible to outliers in the data. Furthermore, the Huber loss function combines the advantageous aspects of both the MAE and MSE metrics.
8) Optimization Procedure: The Adam optimizer, a variant of stochastic gradient descent (SGD) optimization is utilized by the optimization procedure in the provided code. During the training phase, the optimizer works to minimize the Huber loss function.
9) Number of Iterations: The variable that determines the training iterations for the ANFIS model is represented as num_epochs and is set to a value of one hundred in this research. This particular variable administers the number of training epochs or iterations across the dataset throughout the optimization procedure.
10) Hyperparameters: The ANFIS system integrates various hyperparameters, including the number of input variables, number of fuzzy rules, and the learning rate. These parameters also play an essential role in determining the model's capacity, learning behavior, and generalization ability.

Overall, the ANFIS model implemented in this research provides a robust framework for modelling complex and intersectional relationships in datasets while retaining interpretability through fuzzy logic inference mechanisms. Moreover, it effectively addresses various computational tasks by integrating the fusion of NN principles with fuzzy logic paradigms.

3.2 Artificial Neural Network (ANN)

An ANN is a model with multiple layers, where neurons are interconnected nodes. In a non-input layer, the output of each neuron depends on the output of the connected neuron in the preceding layer and the associated activation function. The input layer's neurons receive an input parameter vector and transmit this information to the neurons in the subsequent layer. Hidden layers, not externally visible, consist of intricately connected neurons designed for feature extraction and data processing. The complexity of connections allows for sophisticated neural network operations. The output layer may contain one or more neurons responsible for producing classification outcomes or predictable values.

As shown in **Figure 3.3**, the fundamental architecture of the ANN comprising the input data layer, multiple concealed processing layers, and the data output layer. A pivotal factor within the neural network is the weight denoted as W, featuring a unique value for each pair of connected neurons. With the exception of the input layer, every neuron within the network processes the product of the output data from its connected neuron in the lower layer and the corresponding weight. The output value for each neuron is then determined by the activation function after summing all these products. The essence of neural network training primarily involves iteratively adjusting the weight parameters within the network to achieve optimality [102].

The ANN functions as a machine learning system, requiring constant monitoring, being trained and exposed to extensive datasets to realize practical application effectiveness. Currently, the Back Propagation algorithm (commonly referred to as the BP algorithm) stands as the most widely employed neural network training algorithm [103]. The operational steps of the BP algorithm entail: initially, utilizing data from the training dataset as the neural network input, conducting forward propagation to compute the network's output values. Subsequently, a mean square

cost function is defined to minimize the error between the actual output and the expected output value of the training dataset. The weight parameters are then systematically adjusted by layers via the Gradient Descent Algorithm [104], culminating in the completion of the initial training phase for the neural network. A neural network that is extensively trained with several repetitions is deemed convergent when the error between its output values and the expected values stabilizes within a narrow, consistent range. Once this stable condition is achieved, the neural network is considered reliable and can be effectively applied in practical scenarios.

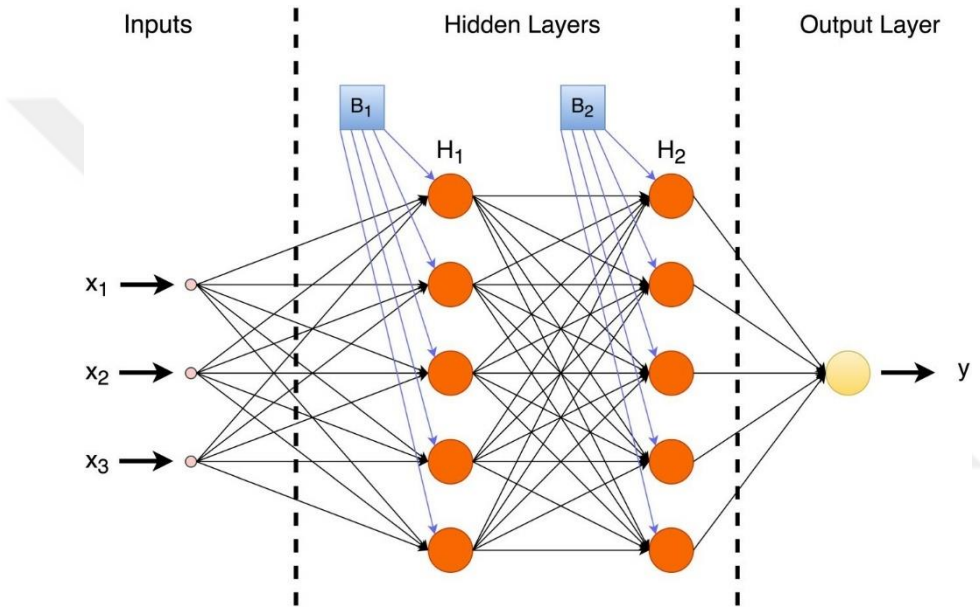
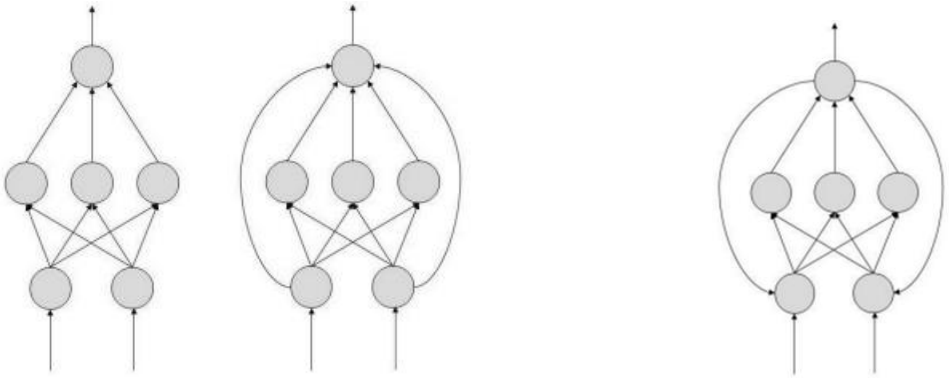


Figure 3.3 ANN architecture [105]



(a) (b)

Figure 3.4 (a) The architecture of Feed Forward and (b) for Feed Back Neural Network topologies systems [106].

The evolution of the ANN, from their conceptualization to the current era marked by rapid advancement, can be delineated into three key stages. The initial stage introduced the perceptron model as the prototype for neural networks, featuring linear connections among neurons. However, this model exhibited minimal practical application value. Following the introduction of nonlinear connections, the pivotal invention of the back propagation algorithm laid the groundwork for neural network training. As the 21st century unfolded, the predominant focus shifted towards deep learning, reshaping the landscape of neural network concepts and significantly enhancing their practical applications. Within the realm of machine learning, the perceptron emerged as a supervised learning algorithm specifically designed for binary classification, determining whether a given numeric vector input belongs to a particular function class [107]. Functioning as a linear classifier, the perceptron employs a classification algorithm grounded in the linear prediction function, amalgamating the weight set and the feature vector for predictions. Originating in the 1950s, the perceptron algorithm stands as one of the earliest implementations of ANNs, realized through specialized hardware.

Two primary ANN topologies are Feed Forward and Feedback. In Feed Forward systems, feedback loops are absent, as each unit transmits data to a different receiver without receiving any in return. Inputs and outputs remain fixed, with each unit obtaining input information from units on its left. These inputs are then reproduced by the weight of each connection, yielding output results associated with the weights of these connections. Applications of this method include pattern generation, identification, and classification. Feed Forward systems are employed in network applications where the desired outcome is already known, such as in computer vision applications [108]. Several viable uses are built upon this foundation. Conversely, in Feedback ANN systems, content-addressable memories are utilized. Learning in neural networks with a feedback process involves comparing the network's output with the desired and expected output. The disparity between these two outputs is leveraged to adjust and alter the connections weights

between network units. The architecture of both the Feed Forward and the Feed Back ANN systems is illustrated in **Figure 3.4**.

This research employs a multilayer perceptron (MLP), a type of ANN utilizing a feedforward architecture, specifically tailored for regression tasks. The characteristics of this ANN for EAC prediction are detailed below:

1) Architecture: The structure of the ANN comprises a structure of an initial input layer (which is succeeded by two hidden layers) and culminates in an output layer. The input layer contains twelve neurons. Each subsequent hidden layer comprises eight neurons that assist in discerning complex patterns within the input data. In contrast, the output layer consists of a solitary neuron designed to generate continuous values particularly suited for regression tasks.
2) Activation Functions: ReLU (Rectified Linear Unit) activation functions are utilized by the hidden layers of the network to incorporate non-linearity and simplify the learning process of the intersecting and complex relationships within the input data. Given its simplicity and effectiveness in addressing the vanishing gradient issues, ReLU activation is preferred to other activation function mechanisms. Conversely, the system employs a linear activation function for the output layer — which is optimal for regression tasks since it creates continuous output values.
3) Optimization Algorithm: The Adam optimizer, a modified form of stochastic gradient descent (SGD), is utilized by the system to improve the model's parameters throughout the training phase. The Adam optimizer adapts the learning rates of each parameter according to their gradients, resulting in improved convergence speed and overall training effectiveness.
4) Loss Function: To measure the difference between the anticipated and actual target values, the MSE loss function is used during the training phase. The model aims to generate predictions that closely estimate the ground truth target values by minimizing the MSE loss. This, in turn, enhances the forecasting accuracy for the regression tasks.
5) Training Procedure: By considering the gradients of the loss function involving the model's parameters, the backpropagation algorithm

iteratively adjusts the weights and biases of the neurons during the training phase. Based on its sequential processes, the training procedure includes feeding the training data to the system through the network, calculating the loss, and adjusting the parameters to minimize the loss. This enhances the model's accuracy and predictability.

6) Evaluation Metrics: The model's performance is evaluated using various evaluation criteria, including (but is not limited to): MAE, R-Value, MSE, and Root Mean Squared Error (RMSE). These indicators provide vital insights related to the model's ability to generalize, and the model's capacity to produce accurate predictions of target values.

Generally, the ANN architecture is accurately designed to detect complex patterns within the data and generate more precise predictions for the regression tasks. The selection of activation functions, optimization algorithms, and evaluation metrics are crucial when optimizing the model's performance and ensuring effective training and convergence.

Table 3.1 presents a list of the benefits and drawbacks of ANN and ANFIS methods. It highlights how the suggested hybrid AOA/FFA-ANN/ANFIS addresses certain limitations of these approaches by outlining the advantages of the proposed model. Furthermore, the drawbacks of the hybrid model are also acknowledged, thus providing valuable insights for researchers and practitioners considering its application.

Table 3.1 Advantages and disadvantages of the ANN, ANFIS, and hybrid AOA/ FFA -ANN/ANFIS

Model	Benefits	Drawbacks	Ref.
ANN	- High-level Architecture: The user can build the model with less formal statistical information about the model's architecture. - High Performance on complex problems: The model can detect complex relationships between dependent and independent variables. - Interaction Detection: The model can detect the possible interactions amongst independent variables.	- Black Box Nature: It is hard to explain the model's predictions or behaviors. - High Computational Cost: requires powerful computation resources like Graphics processing units to run especially for large datasets. - Overfitting Risk: The model risks memorizing the data instead of learning the patterns.	[109], [110], [111], [112]

Table 3.1 Advantages and disadvantages of the ANN, ANFIS, and hybrid AOA/ FFA -ANN/ANFIS (Continued)

Model	Benefits	Drawbacks	Ref.
ANN	- Multiple Training Algorithms: The model can provide access to various training algorithms. - Nonlinear Modelling: The model can effectively capture nonlinearities between dependent and independent variables. - Easy to build: The model is easy to build and requires only basic knowledge of machine learning. - Parallel Processing: The model can perform multiple tasks in parallel, which enhances the overall processing speed. - Distributed Memory: Information is stored across the network, ensuring continued functionality despite localized data loss.	- Empirical Development: The model lacks clear guidelines for determining the best network structures and training algorithms. - Reproducibility: Sharing and reproducing existing models is challenging. - Uncertain Confidence Intervals: Confidence intervals for predictions are difficult to obtain.	
ANFIS	- Combines ANN and Fuzzy Logic: The model merges ANN's learning ability with fuzzy logic to improve rule generation and parameter optimization. - Fast Convergence: The model can converge faster during the training phase. - High Tolerance to Uncertainty: The model manages uncertain data effectively. - Excellent Generalization: The model provides strong interpolation, extrapolation, and generalization capabilities.	- Black Box Nature: it is hard to explain the model's predictions or behaviors. - Sensitivity to Initial Rules: The model's performance is sensitive to the initial number of fuzzy rules (clusters). - Increased Computational Complexity: More Computational Resources are required when the number of fuzzy rules grows. - Input Limitations: The model is limited by the number of inputs that it can manage effectively.	[113], [110], [114], [115].
Hybrid Model	- Enhanced Feature Selection: AOA/ FFA is better at effectively selecting the most relevant features, thus improving model performance.	- Black Box Nature: Combining multiple algorithms can make the	

Table 3.1 Advantages and disadvantages of the ANN, ANFIS, and hybrid AOA/ FFA -ANN/ANFIS (Continued)

Model	Benefits	Drawbacks	Ref.
Hybrid Model (AOA/FFA + ANN / ANFIS)	- Reduced Dimensionality: The model uses fewer input variables, making the model simpler and faster. - Improved Accuracy: The model selects optimal features, which in turn, often enhances the model's accuracy. - Better Generalization: The model reduces overfitting by eliminating irrelevant or redundant features. - Increased Efficiency: The integration of AOA/FFA improves the training process, saving computational resources. - Adaptability: The model can be designed to adapt to diverse types of data and applications.	model harder to interpret and understand. - Complexity of Integration: Combining AOA/ FFA with ANN/ANFIS can be complex and requires advanced expertise. - Computational Overhead: The initial feature selection phase adds more computational complexity. -Algorithm Sensitivity: Performance can be sensitive to AOA/FFA parameters. - Resource Intensive: Especially for large datasets, computational resource consumption may increase. Although it is more efficient overall, the hybrid approach still requires substantial computational resources for large datasets.	

3.3 Variables Input Section

The performance of machine learning models is significantly influenced by the selection of input variables. As the dimensionality of data increases, more training samples are required for model reliability. However, due to limited training sample availability, dimension reduction has become a crucial step. It is important to ensure that no essential information is lost during data dimension reduction, in order to train a model properly. While technical advancements have enabled vast amounts

of data collection, the primary challenge remains the extraction of useful knowledge from such data. The number of correlated input variables included in the training process can impact prediction performance, so it is important to significantly reduce dimensionality to ensure model training accuracy and reliability.

3.4 Optimization Algorithms:

In recent years, numerical optimization problems have become more intricate, leading to the need for more efficient strategies to solve them. In various fields, such as engineering and data mining, achieving optimal results while minimizing costs and maintaining accuracy can involve searching through a large number of available options. However, exploring sub-optimal areas can be time-consuming, so efficient approaches are needed to quickly locate the best solution. Optimization algorithms face significant challenges due to the complex structure and highly non-convex terrain of the search space associated with these problems [116]. Traditional methods based on calculus principles and exhaustive search are not as effective as heuristic methods in terms of computational resources. Metaheuristic algorithms, also known as non-deterministic approaches, have proven to be very effective in solving complex real-world optimization problems [117, 118]. The application of metaheuristic research is diverse and spans across various fields such as motif discovery, drug design, finance, feature selection, engineering, agriculture, medicine, and economics. The use of metaheuristic algorithms has proven successful in solving a wide range of practical optimization problems. Well-known metaheuristic algorithms that have produced remarkable results include simulated annealing (SA), PSO, ant colony optimization, GAs, and artificial bee colony [119]. The Firefly and Archimedes optimization methods will be utilized in this research.

3.5 Archimedes Optimization Algorithm:

Hashim et al. [119] developed the AOA based on the Archimedes principle, a fundamental physics law concerning objects submerged in fluids. This principle states that an object generates buoyancy, or an upward force, equal to the weight of the fluid it displaces when submerged. In the AOA, submerged objects serve as individuals in the population or potential solutions. The first step of the technique is to initialize a population of objects with randomly assigned initial positions

within the problem's search space. Subsequently, the fitness function for each object is computed. Throughout the iterative process, AOA dynamically updates the volumes and densities of the objects. The accelerations of the objects are modified based on their collisions with neighboring objects [120]. The following formula is used for all object initialization processes:

$$O_i = l_i + rand \times (u_i - l_i), \forall i \ \in \{1,2,3, \ldots., N\} \quad (3.6)$$

N is the number of objects. The lower limit (li) and upper limit (ui) of the ith object are defined. Each object's density and volume can be configured as follows:

$$Den_i = rand \quad (3.7)$$

$$Vol_i = rand \quad (3.8)$$

In this case, the D-dimensional vector "rand" has values between 0 and 1. Each object's acceleration can be calculated using the following formula:

$$ACC_i = lb_i + rand(0,1) \times (ub_i - lb_i), \forall i \ \in \{1,2,3, \ldots., N\} \quad (3.9)$$

The object with the highest fitness is selected as x_{best}, vol_{best}, den_{best}, and acc_{best} after considering the initial fitness function.

The following formula is used to update the density and volume of the ith item:

$$Vol_i^{(t+1)} = Vol_i^{\,t} + rand \times \left(Vol_{best} - Vol_i^{\,t}\right) \quad (3.10)$$

$$Den_i^{(t+1)} = Den_i^{\,t} + rand \times \left(Den_{best} - Den_i^{\,t}\right) \quad (3.11)$$

"t" stands for the iteration that is currently in progress, while "rand" is a random number in this algorithm. The items collide at first, and then each tries to come to a state of balance. Using a transfer operator to facilitate the transition from exploration to exploitation phases, AOA demonstrates this process. The transfer operator formula is represented as follows:

$$TF = \exp\left\{\frac{t - t_{max}}{t_{max}}\right\} \quad (3.12)$$

The symbol "t_{max}" in this equation indicates the maximum number of iterations. With each iteration, the TF value steadily increases until it reaches unity. Another component that helps AOA make the shift from global to local search is the density-lowering factor, which has the following formula:

$$d^{t+1} = \exp\left\{\frac{t - t_{max}}{t_{max}}\right\} - \left(\frac{t}{t_{max}}\right) \tag{3.13}$$

The value of d^{t+1} decreases with time. Setting this variable correctly helps in attaining a balance between exploring and exploiting. Exploration occurs when objects collide. The collision takes place when the transfer operator is at 0.5. By selecting a random material (mr) using this procedure, the acceleration of the i^{th} object in the subsequent iteration (t + 1) is updated as follows:

$$ACC^{t+1} = \frac{Den_{mr} + Vol_{mr} \times ACC_{mr}}{Den_i{}^{(t+1)} \times Vol_i{}^{(t+1)}} \tag{3.14}$$

The density, volume, and acceleration of the randomly selected material (mr) are represented in this section by the variables den_{mr}, vol_{mr}, and acc_{mr}. When there are no collisions between objects, the AOA enters its exploitation phase, which begins when the value of the transfer operator exceeds 0.5. The formula below determines the ith object's acceleration during the exploitation phase:

$$ACC_i^{t+1} = \frac{Den_{best} + Vol_{best} \times ACC_{best}}{Den_i{}^{(t+1)} \times Vol_i{}^{(t+1)}} \tag{3.15}$$

The best object density, volume, and acceleration are denoted, respectively, by den_{best}, vol_{best}, and acc_{best} in this equation. Since it establishes the proportion of change for every particle, normalizing the acceleration of each particle is essential. The following formula is an expression for normalized acceleration:

$$ACC_{i-norm}^{t+1} = u \times \frac{ACC_i{}^{t+1} - \min\{ACC\}}{\max\{ACC\} - \min\{ACC\}} + l \tag{3.16}$$

l and u denote the normalization ranges in this formula, which are set to 0.1 and 0.9, respectively. The acceleration value is large when an object is distant from the global optimum, indicating the exploration phase. If not, the exploitation phase is applied; in this case, the exploration phase is conducted. During the exploration phase, the following formula is used to update the position of the ith particle:

$$x_i{}^{t+1} = x_i{}^t + C_1 \times randACC_{i-norm}^{t+1} \times d \times (x_{rand} - x_i{}^t) \tag{3.17}$$

As an alternative, the following diagram shows how the positions of the particles are updated during the exploitation phase:

$$x_i{}^{t+1} = x_{best}{}^t + F \times C_2 \times randACC_{i-norm}^{t+1} \times d \times (T \times x_{best} - x_i{}^t) \tag{3.18}$$

Here, T is a variable that depends on the transmission function (T = C3 × TF), whereas C3 is a constant. The user sets C1 and C2. The best particle's position is indicated by xbest, and the flag F is used to change the particle's direction of travel. The following formula is used to determine the value of F:

$$F = \begin{cases} +1 \; if \; P \leq 0.5 \\ -1 \; if \; P > 0.5 \end{cases} \tag{3.19}$$

Within this formula, the user arbitrarily assigns the value of P. Subsequently, the fitness function is computed based on the updated positions of the particles, and the optimal solution is recorded.

3.6 FFA (FIREFLY) Optimizer Algorithm

The Firefly algorithm (FFA) was introduced by Yang in 2010 [121]. The algorithm was inspired by the light interactions among fireflies. It exemplifies the collective swarm intelligence in which the potential competition and collaboration among members (with simple and low intelligence) results -when united - in a higher level of intelligence that cannot be normally achieved by any individual member/element.

The formulation of the FFA is established upon a set of rules:

1) Fireflies lack gender distinctions, exhibiting attractive and transparent movements irrespective of their sex. Typically, a firefly absorbs light emitted by other fireflies.
2) A firefly's allure correlates directly with its brightness, which diminishes with distance from other fireflies. Without a more enticing firefly, it resorts to random movement.
3) The objective function's value dictates the luminosity of a firefly.

The foraging behavior and social interactions observed in fireflies were the main inspiration that drove the creation of the FFA. It was observed that the fireflies move in seemingly random patterns while seeking food, with individual flies emitting brighter light to tell the group of the existence of better food resources and attract close by fireflies. As the distance between fireflies increases, the intensity of attraction diminishes. This indicates an inverse relationship between distance, velocity, and attraction. This algorithm draws on two fundamental factors: variations in light intensity and movement towards brighter fireflies. Light intensity

corresponds to the cost function value, particularly in minimization problems, where brighter fireflies attract those emitting lesser light. The symbol 'n' denotes the number of fireflies, 'xi' represents the location of the 'i-th' particle, and 'f(xi)' signifies the cost function value associated with it. Therefore, the brightness of each firefly aligns with the respective cost function value:

$$Ii = f(x_i) \qquad 1 < i < n \tag{3.20}$$

Every firefly possesses a luminous feature that indicates its strength, with this trait being a relative value influenced by alterations in the distance between firefly i and firefly j. The attraction function is expressed below as follows:

$$\beta r = \beta_{\circ} * e^{-\gamma r2} \tag{3.21}$$

The degree of attraction, represented as $\beta_{\circ}$ at r=0, and γ which indicates the coefficient of light attraction, and outlines the parameters. The calculation of the movement from firefly i located at xi towards firefly j located at xj (lighter one) is expressed as follows:

$$x_i(t+1) = x_i(t) + \beta_r(x_j - x_i) \tag{3.22}$$

In the context of the FFA, the flowchart can be depicted as follows [122]:

1) Initialization of fireflies entails determining and setting the initial population.
2) Estimating the cost function for everyone in the population based on its whereabouts.
3) Light intensity is randomly assigned to each member within the population.
4) Identifying the best (brightest) member is achieved by calculating the cost function for the entire population.
5) Other individuals in the population move towards the best individual, resetting their light intensity accordingly.
6) Iteration conditions are checked; if provided, the algorithm progresses to the next stage, otherwise, it returns to step 4.
7) The algorithm concludes at when the process is over.

Various factors (such as the attraction function, the initial population size, and the attraction coefficient) drastically impact the effectiveness of the FFA. An increased speed of attraction for population members toward the brightest member means a higher value for the light attraction coefficient.

3.7 Model Accuracy:

Various numerical indicators, such as RMSE, MAE, mean relative error (MRE), Nash-Sutcliffe coefficient (NSE), scatter index (SI), and correlation coefficient (R) [17], are used to evaluate predictive models, which allow for an evaluation of the absolute error (the closer to zero the better) and the quality of the best model (the closer to one the better). The optimal model can thus be more accurately justified using the best input combination. The expressions for these indicators are as follows:

$$\text{MAE}=\frac{1}{n}\sum_{i=1}^{n}\left|Y_i-\hat{Y}_i\right| \tag{3.23}$$

$$\text{MSE}=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2 \tag{3.24}$$

$$\text{R}=\frac{\sum_{i=1}^{n}(Y_i-\bar{Y})\sum_{i=1}^{n}\left(\hat{Y}_i-\bar{\bar{Y}}\right)}{\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2\sum_{i=1}^{n}\left(\hat{Y}_i-\bar{\hat{Y}}_i\right)^2}} \tag{3.25}$$

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2} \tag{3.26}$$

$$\text{Symmetric Mean Absolute percentage error (SMAPE)}=\frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i-\hat{Y}_i|}{(|Y_i|+|\hat{Y}_i|)/2} \tag{3.27}$$

$$\text{MRE}=\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i-\hat{Y}_i}{|Y_i|} \tag{3.28}$$

Were:

Y_i is the actual value, $\hat{Y}_i$ is the predicted value, $\bar{Y}$ is the average of actual value, and $\bar{\bar{Y}}$ is the average of predicted value. n is the number of data. Python was used to solve every equation after examining the expected value with the same software.

3.8 Modeling Procedure Phase

3.8.1 Programming Information and used Codes

When diving into programming, the selection of a programming language is a critical initial task, especially for beginners. The chosen language plays a pivotal role in shaping one's journey toward becoming a programmer. Python, in this context, stands out as a versatile, general-purpose programming language supporting multiple paradigms. Recognized for its beginner-friendly nature, Python minimizes coding complexity and facilitates a smooth entry into programming. It

boasts a high-level structure that accomplishes tasks with minimal lines of code compared to other languages. Python's user-friendly coding style and extensive standard library make it an ideal starting point. Key features include simplicity, free and open-source usage, high-level programming, platform independence, portability, dynamic typing, support for both procedural and object-oriented programming, interpretability, extensibility, and an extensive library [123]. Python's development began in the late 1980s under Guido van Rossum, and its first release was in 1991 as Python 0.9.0. The language's dynamism, showcased by its dynamic typing, eliminates the need for explicit data type declarations. Python 2.0 was released in 2000, followed by Python 3.0 in 2008, introducing changes that weren't completely backward-compatible with earlier versions [124]. In terms of popularity, Python has consistently ranked as the number one programming language according to Popularity of Programming Language in 2022 [125]. Beyond coding, Python simplifies the creation of Graphical User Interfaces (GUIs), allowing users to control software programs through a point-and-click interface, making it accessible for non-programmers. This flexibility extends to developing GUI applications that can be used within Python or as standalone computer or web applications. The screenshot of the python code for AOA-ANN, AOA-ANFIS, FFA-ANN and FFA-ANFIS is shown in the Figures 3.5-3.8 respectively.

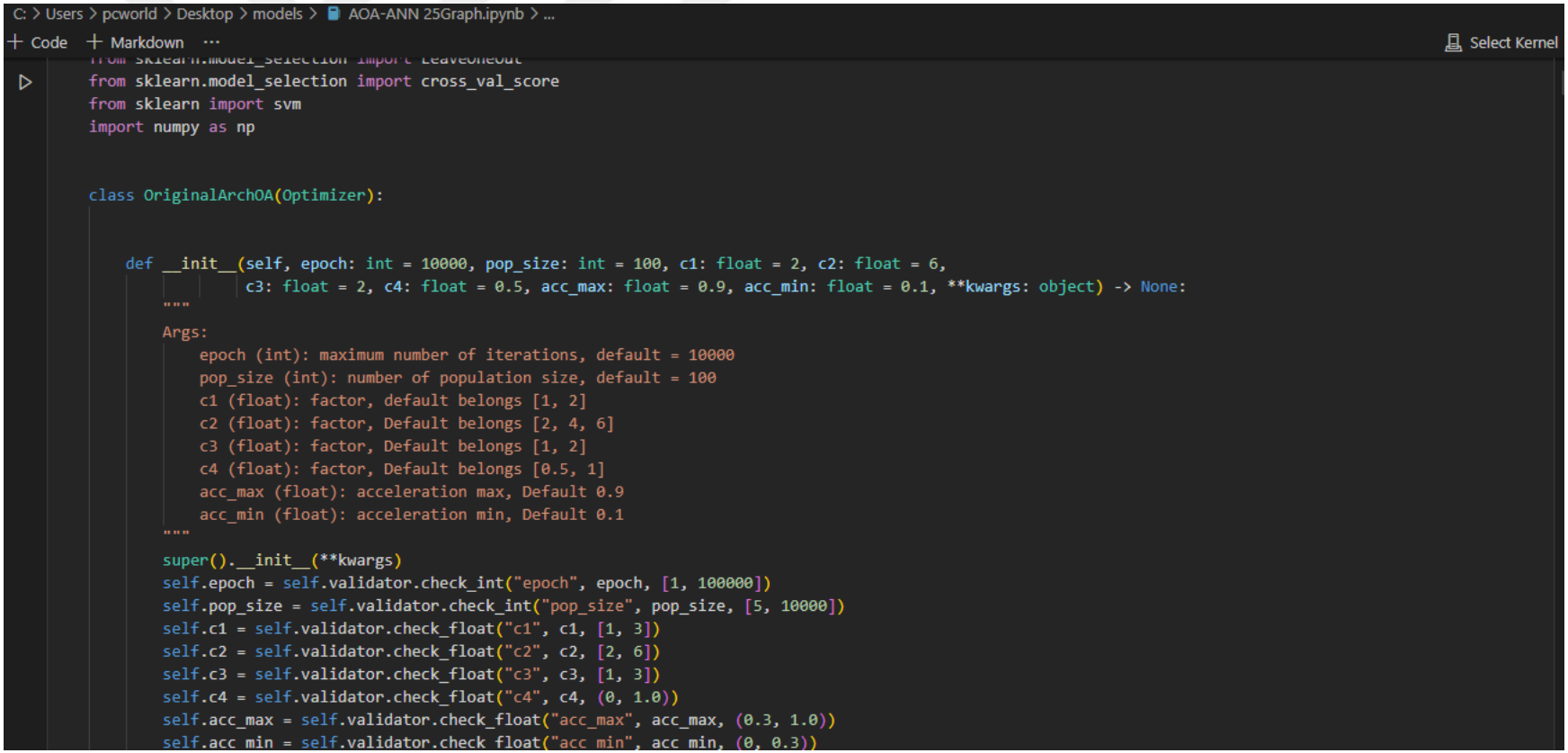


Figure 3.5*:* The AOA-ANN screenshot code.

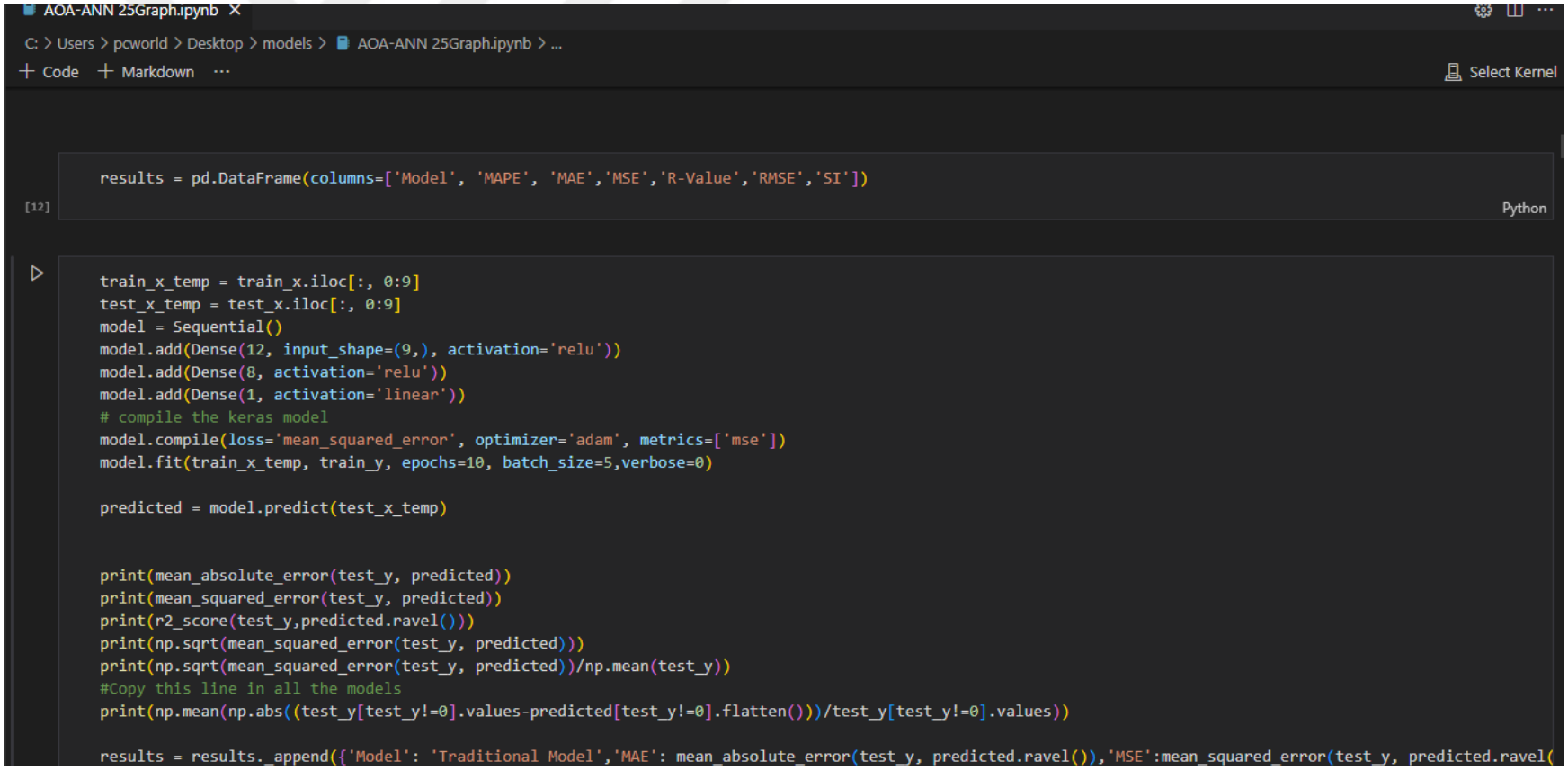


Figure 3.6: The AOA-ANN screenshot code.

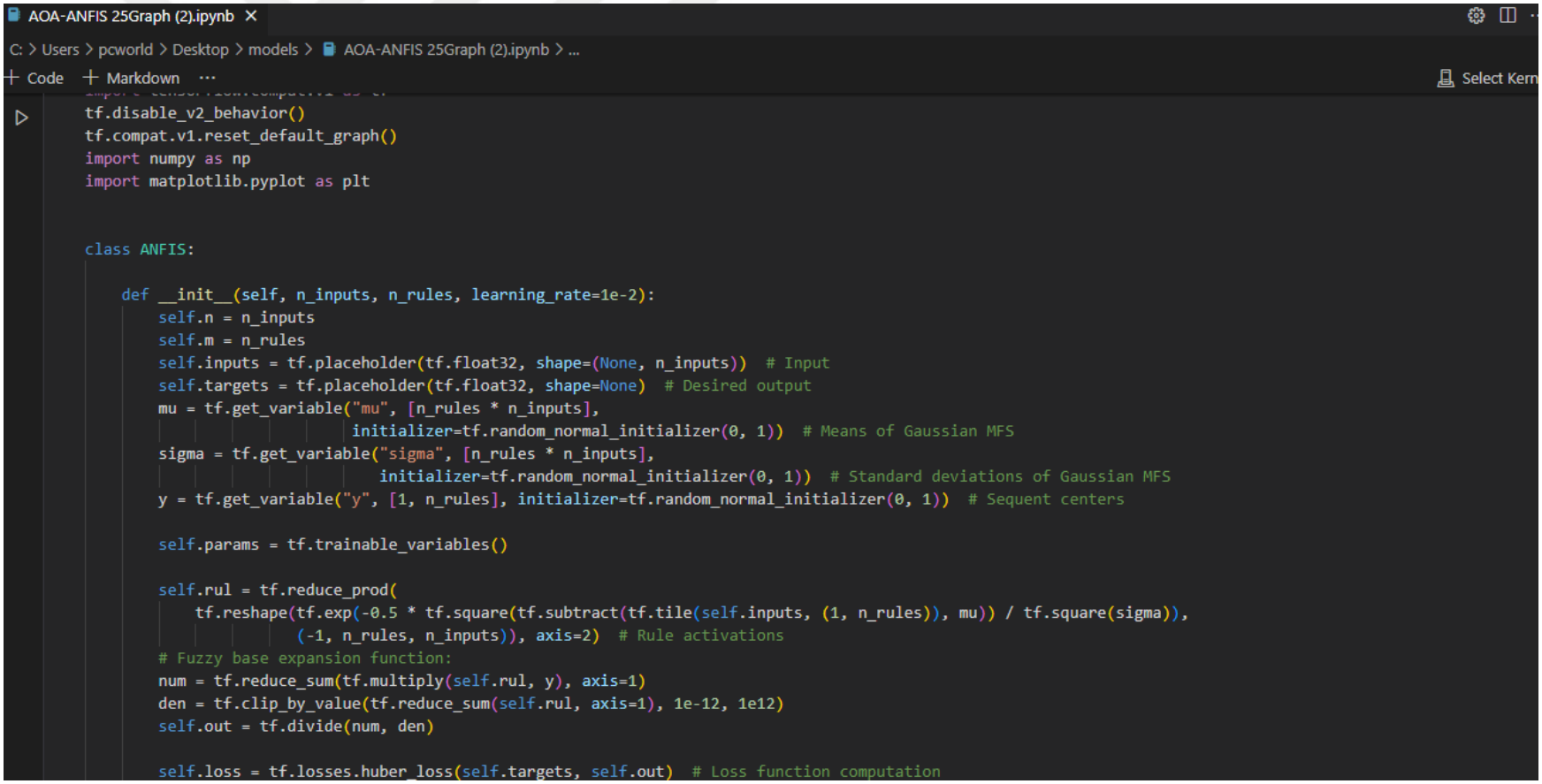


Figure 3.7: The AOA-ANFIS screenshot code.

AOA-ANFIS 25Graph (2).ipynb | FFA_ANN_Optimized.ipynb

C: > Users > pcworld > Desktop > models > FFA_ANN_Optimized.ipynb

+ Code + Markdown ··· Select Kernel

```python
class FireflyAlgorithm():
    def __init__(self, function, dim, population_size=30, max_evaluations=10, alpha=1, betamin=0.5, gamma=1):
        self.D = dim
        self.NP = population_size
        self.nFES = max_evaluations
        self.alpha = alpha
        self.betamin = betamin
        self.gamma = gamma
        self.Fireflies = np.random.rand(self.NP, self.D)
        self.Fitness = np.zeros(self.NP)
        self.I = np.zeros(self.NP)
        self.LB = 0
        self.UB = 1
        self.Fun = function
        self.evaluations = 0

    def move_ffa(self):
        for i in range(self.NP):
            for j in range(self.NP):
                if self.I[i] > self.I[j]:
                    r = np.sqrt(np.sum((self.Fireflies[i] - self.Fireflies[j])**2))
                    beta = self.betamin + (1 - self.betamin) * np.exp(-self.gamma * r ** 2)
                    self.Fireflies[i] += beta * (self.Fireflies[j] - self.Fireflies[i])
                    self.Fireflies[i] += self.alpha * (np.random.rand(self.D) - 0.5)
                    self.Fireflies[i] = np.clip(self.Fireflies[i], self.LB, self.UB)

    def evaluate_fireflies(self):
        # Parallel evaluation of the fireflies
        num_cores = multiprocessing.cpu_count()
        self.Fitness = Parallel(n_jobs=num_cores)(delayed(self.Fun)(firefly) for firefly in self.Fireflies)
```

Figure 3.8*:* The FFA-ANN screenshot code.

3.8.2 The Work Methodology

1. **For Single AFNIS and ANN Models**

Figure 3.9 depicts the operation of a single AFNIS model using the AFNIS methodology. Initially, the input and target information are gathered, followed by the training of the AFNIS system with 75 percent of the data, and ultimately, the remaining 25 percent is utilized for conducting tests. Then, numerical indicators are computed.



Figure 3.9: AFNIS and ANN models methodology

2. **For Hybrid Models**

The operation of the suggested system is depicted in **Figure 3.10-3.13**; Improvement strategies have been used in various AI applications. A large portion of these applications fall into four principal classifications:

1) the selection of features,
2) the training of neural networks,
3) the improvement of SVM, and
4) the applications of clustering.

Feature selection is one of the important processes in machine learning and data mining. The goal of feature selection is to reduce the number of features, select the most representative ones and to eliminate redundant, noisy and irrelevant features. The problem of searching for best set of features is considered as complex and difficult problem due to the extremely large search space when the number of features is large.

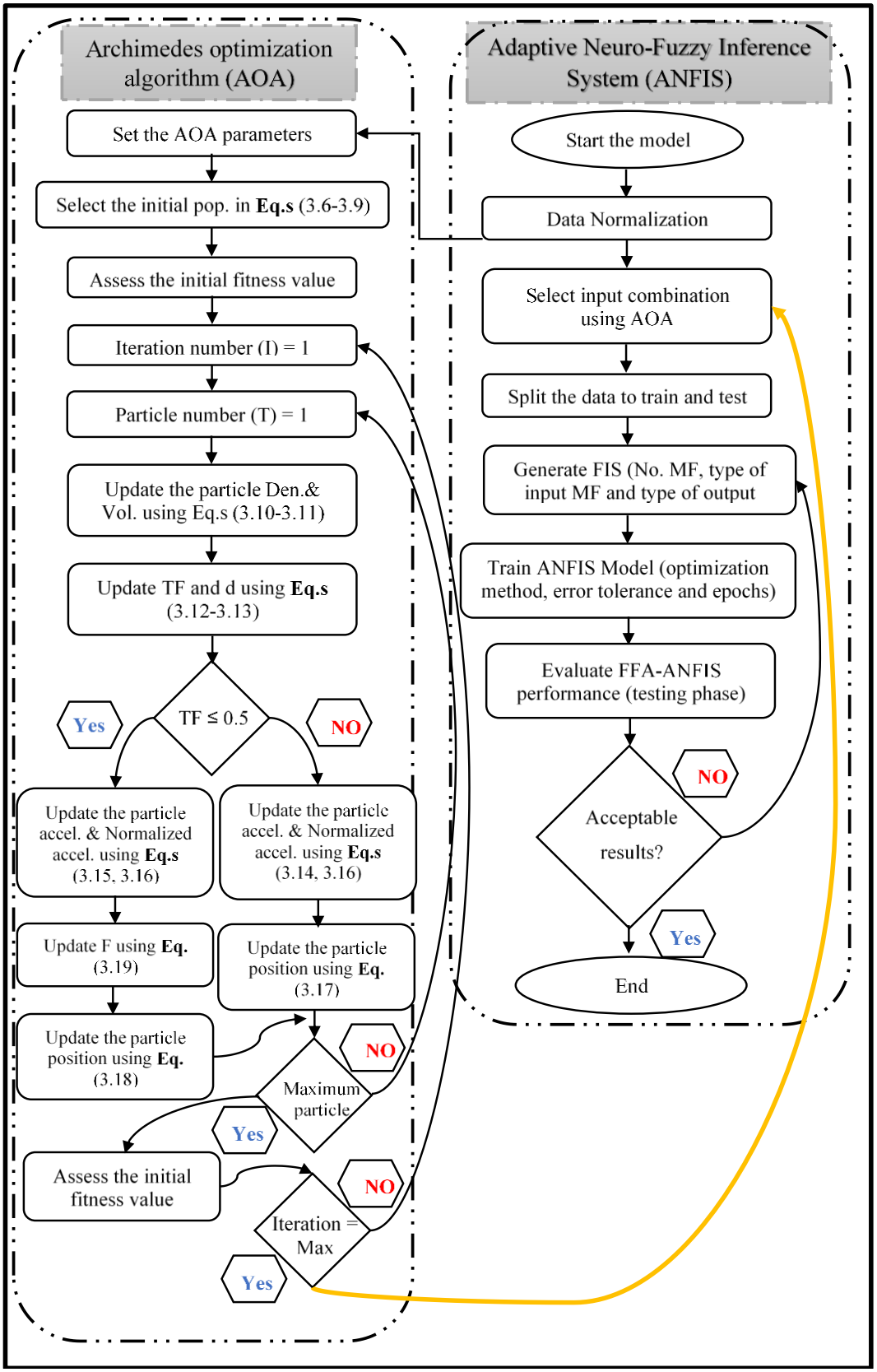


Figure 3.10: Hybrid AOA- ANFIS model Methodology.

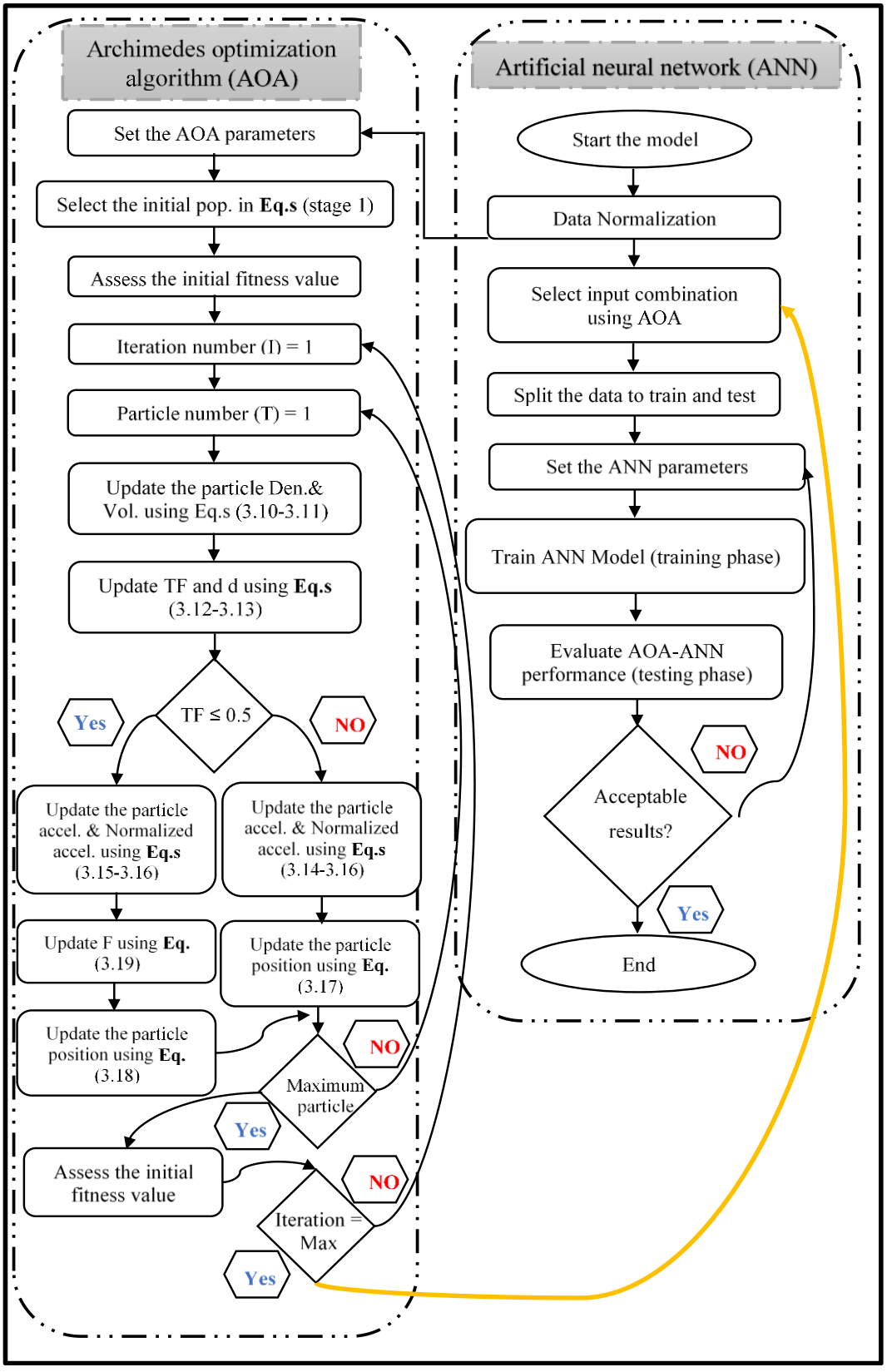


Figure 3.11: Hybrid AOA- ANN model Methodology.

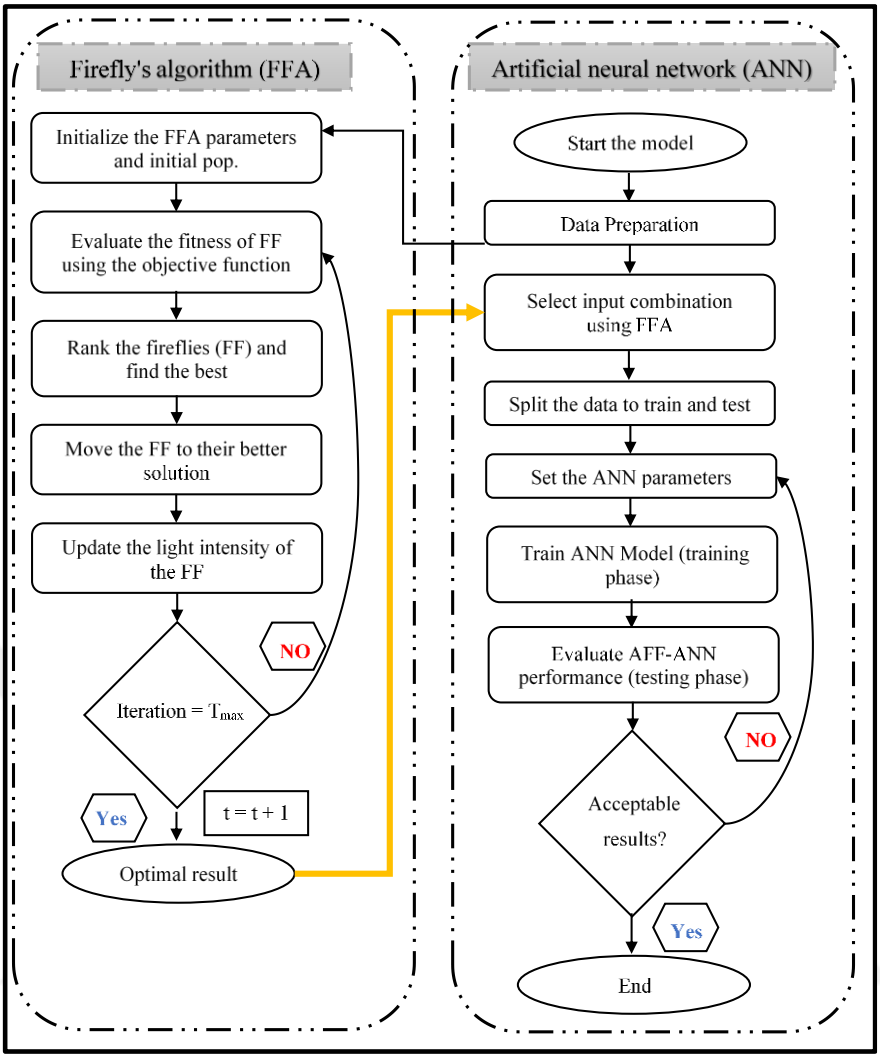


Figure 3.12: Hybrid FFA- ANN model Methodology.

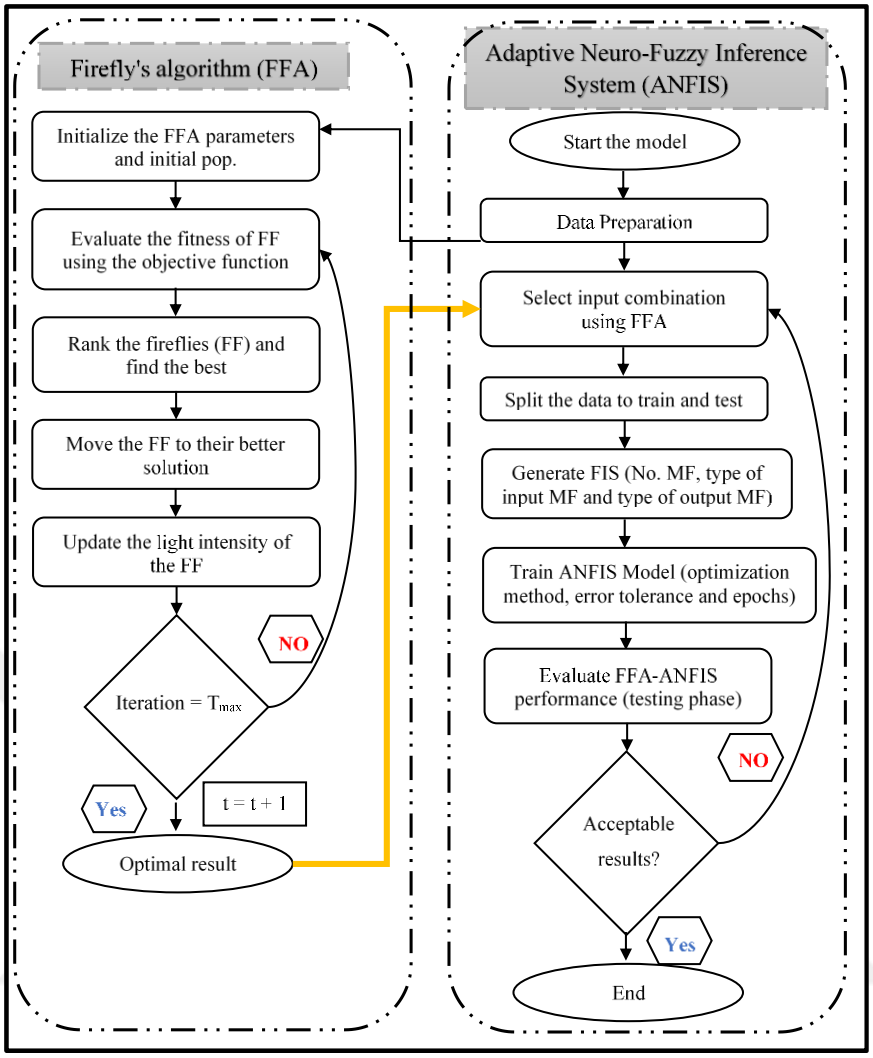


Figure 3.13: Hybrid FFA- ANFIS model Methodology.

3.9 Influence Variables

The building projects examined in this research were built in northern Taiwan between 2000 and 2007, all of which were completed by a single construction company located in Taipei. The height of buildings ranges from nine to seventeen stories, including basement levels. The contract values of the buildings surveyed in this study varied between NT$80 million and NT$1.1 billion. The buildings' total floor space ranged from 2094m^2 to 31,780m2. The construction periods vary between 15 and 63 months. Relevant data set information is provided in **Table 3.2**. Training and testing sets comprise historical data sets which equal 306 data sets (the training set contains 229 data points, whereas the testing set has 77 data points). This data size can be deemed appropriate given that Cheng et al. utilized the same data set for their research [17]. Besides, there are studies that use a similar number

of data points in their studies for predicting the EAC. For instance, Huang & Cheng utilized 289 training data and 17 testing data from 15 projects [19]; and Narbaev & Demarco utilized training cases from nine projects [28]. Similarly, AlHares employed 132 data points from ten projects [30]. Finally, Kamoona utilized 15 projects for a total of 174 data points [16].

Table 3.2 Information of historical data [17].

Project Name	Total Area (m^2)	Underground Floors	Ground Floors	Buildings	Start Date	Finish Date	Duration (days)	Contract Amount (NTD)	Prediction Periods
A	12,622	2	9	1	2003/12/1	2005/8/22	630	289,992,000	29
B	4,919	3	11	1	2003/12/13	2005/11/10	698	149,300,000	24
C	19,205	5	8	1	2000/5/20	2002/5/19	729	332,800,000	20
D	5,358	3	9	1	2000/11/15	2002/11/14	729	199,600,000	25
E	27,468	2	11	3	1999/12/16	2001/12/3	718	1,142,148,388	26
F	31,797	2	9	4	2001/7/4	2003/3/31	635	530,000,000	20-
G	7,707	2	14	1	2001/11/24	2003/10/20	695	153,500,000	22
H	10,087	3	14	1	2002/6/18	2004/7/6	749	216,000,000	27
I	3,479	1	10	1	2003/6/2	2004/9/30	486	85,714,286	18
J	6,352	4	11	1	2004/3/5	2006/2/18	715	202,241,810	31
K	4,774	2	11	1	2004/2/21	2006/2/20	730	145,377,589	27
L	7,289	2	8	1	2005/6/15	2006/9/15	457	190,844,707	20
M	3,094	2	7	1	2005/10/1	2007/2/28	515	102,500,000	17
Total				306					
Subtotal (Training Cases)				229					

The monthly collection of cost records from contractors for each data set and the identification of 10 EAC calculation parameters. In this study, the historical data

was taken from [17,18,19] studies. In these studies, 10 influential aspects of the EAC prediction problem are discovered. **Table 3.3** shows the variables that were analyzed [18].

Table 3.3 Factors that exert an impact [18].

Factor	Factor	Index	Definition
F1	Construction duration	Construction progress percentage	Duration to date/revised contract duration
F2	Actual cost	AC_p	Actual cost/budget at completion
F3	Planned cost	EV_p	Earned value/budget at completion
F4	Cost management	CPI	Earned value/actual cost
F5	Time management	SPI	Earned value/planned value
F6	Subcontractor management	SBI	Subcontractor billed amount/actual cost
F7	Contract payment	OBI	Owner billed amount/earned value
F8	Change order	COI	Revised contract amount/budget at completion
F9	Construction price fluctuation	CCI	Construction material price index of that month/construction material price index of initial stage
F10	Number of rainy days	CEI	(Revised project duration – number of rainy day)/revised project duration

The project completion date is primarily influenced by certain key parameters such as SV, SPI, CV, and CPI, according to the Integrated Program Management Division of the National Defense Industrial Association [16]. Anbari asserts that the value of project management's earnings is predominantly reliant on variations in cost and schedule as well as performance indices, in addition to forecasts of project cost and schedule beyond completion. Therefore, these variables are likely to exert a substantial influence on the project's finalization date [15].

Cheng et al. employ an evolutionary inference model based on support vector machine to analyze the EAC of construction projects. Construction progress (%), CPI, ACp, EVp (earned value percentage), SBI, SPI, COI, OBI, and CCI were utilized for forecasting. Furthermore, the 269 periods gathered from 11 projects were included in the database. Using the ESIM performance evaluation

module, it was possible to determine the proportion of potential costs for each project. The estimated value is known as the "Expected output" [17].

Kamoona proposed a new approach for estimating the EAC, utilizing a deep neural network (DNN) model, which was compared against a SVR model. The study involved fifteen projects located in Iraq, where historical data was used to evaluate the performance of the DNN and SVR models. The research also integrated two input optimization algorithms, including a GA and a BF, with both the proposed and comparable predictive models. The input optimization strategies aimed to enhance the prediction accuracy, and several criteria were used in the forecasting, such as CV, CPI, SV, SPI, CEI, SBI, CCI, COI, and OBI [16].

Huang and Cheng's research focuses on developing a novel evolutionary fuzzy SVM model that can assess the completion of construction projects based on several parameters. In this study, the selected parameters used for forecasting include construction progress percentage, CPI, ACp, earned value percentage, SBI, SPI, COI, OBI, and CCI. The researchers utilized a dataset of 306 data sets, of which 269 were used for training and 37 for testing the model. The model's effectiveness was evaluated based on its ability to accurately predict the completion of construction projects [19].

The study conducted by Feylizadeh et al. explores the use of a hybrid fuzzy-neural network model for predicting the completion costs of construction projects. The suggested technique incorporates both qualitative and quantitative elements influencing the EAC forecast. The chosen criteria employed during prediction were the experience of the project team, weather conditions, SPI, employer cash status, PV, CPI, budget at completion, and actual costs [86]

AlHares utilized a combination of new intelligent models to predict the construction EAC. They employed the global harmony search (GHS) and extreme learning machine (ELM) algorithms in conjunction with a radial basis function network (RBFN) model. The selected parameters for prediction were project CV, CPI, SV, SPI, SBI, OBI, CPF, and COI [30].

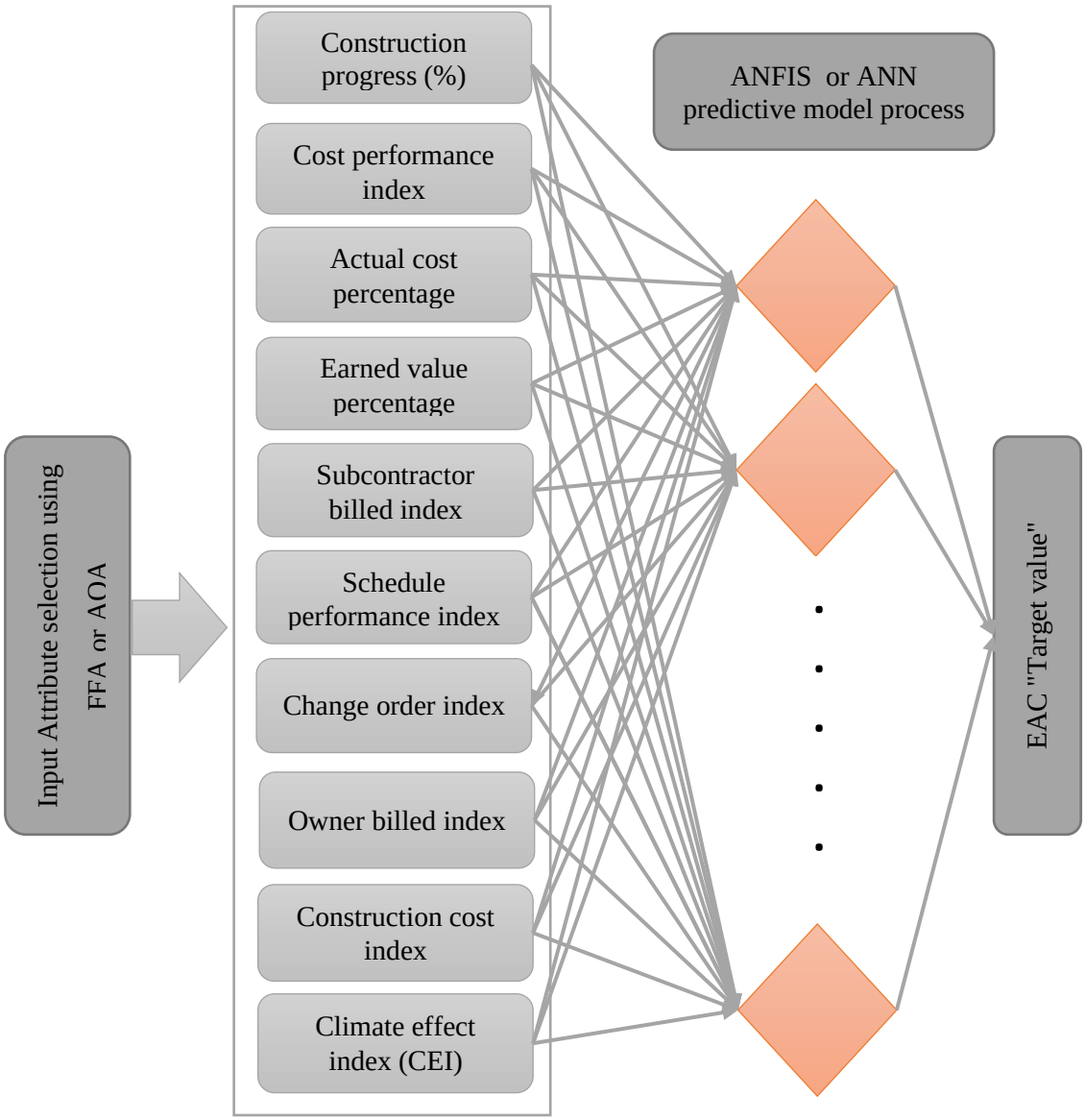


Figure 3.14 The developed EAC predictive model

In EVM, the construction progress (%), ACp, and EVp (earned value percentage) are crucial criteria. They assist you in analyzing the progression of the project, i.e., your performance in terms of time and budget. Variance analysis plays a crucial role in the success of every project that is completed on time and within budget. It involves assessing the schedule and cost variances, which are fundamental components used to determine if a project is behind or ahead of schedule and if it is under or over budget. By analyzing variances, project managers can make informed decisions and take appropriate actions to keep the project on track [126]. Evaluating the condition of a project can be done through schedule and cost variations. It is essential for project managers to keep an eye on these variations and

detect any anomalies. A positive variation in both schedule and cost indicates that the project is on track. Conversely, if both variations are negative, there may be a problem that needs to be addressed through corrective measures. Monitoring variations is crucial to ensure project success and timely delivery [30].

The CPI is determined by the ratio of the average benefit received (BCDP) to the total cost (ACDP) of the project. When the CPI is higher than 1, it indicates that the project's completed deliverables were achieved at a lower cost than originally estimated. If the CPI is below 1, real expenses exceed projected costs. The CPI is an efficient tool for assisting project managers in determining which projects warrant their attention. Example: If there are several projects (A, B, and C) with CPIs of 1.20, 0.75, and 1.05, respectively. Project (B) requires attention however, Project (C) with a CPI of 1.05 is likely to be satisfactory. The director's attention is likely required for the project (A). Its CPI of 1.20 may suggest quality difficulties resulting from significant cost overruns. [9,25,127].

The measurement of schedule deviation is often based on the ratio between the cumulative budgeted cost of produced deliverables and the expected cumulative budgeted cost, known as the SPI. This metric is widely used in project management. In a broader sense, the SPI represents the ratio of actual deliverables to anticipated deliverables. Variations in expenditures are a commonly adopted success indicator, according to practitioners of project management. No other behavior was considered during the analysis. The SPI is calculated by dividing the EV by the PV, indicating the degree of schedule deviation. An SPI value of 1 suggests that the project is on track and delivering more output than planned. When SPI falls below 1, the production of deliverables is behind schedule. For calculating changes in plan timelines, the project performance index (SPI) was employed. The SPI is calculated as the percentage of the Deliverables Budget that was actually created to the deliverables budget that was scheduled. When SPI is 1.0, deliveries are on time; when it is less than 1.0, they are late; and when it is greater than 1.0, they are delivered earlier than anticipated [128,129,130].

Change orders are possible and frequently inevitable during building projects. A change order refers to an alteration made to the original scope of work, which may involve the addition or removal of tasks; therefore, the original contract price and/or

the completion date of your project are changed [131]. Despite the building team's best efforts to avoid them, modification orders are prevalent. There are several causes for transformation. Typically, they arise from unanticipated field circumstances, design modifications, or owner requirements.

Change orders might raise your project's cost. Plan for a 5-10% contingency to cover these expenses in order to prevent going over budget. Consequently, if modifications arise, they will not be over your budget. When a change is necessary, it is important to inform the team as soon as possible. This allows for the problem to be explored and a solution to be devised. Usually, the adjustments will need the architect or engineer to update the project's designs or make revisions to the required permits, after which the owner must approve the added cost work. The project schedule can be affected to varying degrees by a change, with some tasks possibly taking longer to complete or the entire schedule being extended [132].

SBI, which is expressed as subcontractor billing amount / actual cost contractor can refer to either the main contractor or a subcontractor. Material procurement accounts for an average of forty percent of the total cost of manufacturing; consequently, subcontractor management is vital. Therefore, a significant fraction of quality issues is subcontractor related. The formation of a partnership is vital for the success of both parties' businesses. The subcontractor should contribute positively to design, execution, and expense savings. The whole material cost, which includes price and quality, should be emphasized [133].

The OBI is a metric that represents the ratio of the owner billed amount to the EV of a project as specified in the contract payment agreement. Even if they are required to adhere to the timeline, contractors may not be able to perform any work if insufficient funds are available at any time throughout construction. The capacity of construction firms to conduct construction activities with minimum finance expenses is crucial to their profitability. In lump sum projects, contractors are often compensated based on their proven percentage of completion, as well as the agreed-upon income (as specified in the contract) for the finished job [134].

Climate change raises the dangers associated with initiating construction projects. The building industry is affected by environmental and atmospheric changes, and climate change can have both direct and indirect impacts. Changes in climate can

lead to on-site planning challenges, disruptions, increased expenses, worker health issues, fluctuations in material costs, and transportation delays [134]. There is a growing body of evidence that indicates climate change poses significant risks. In order to effectively manage projects, project managers must develop new skills and knowledge to better understand how climate change may impact project safety and success [135].

The CCI is a weighted price index that is calculated using constant material quantities. It takes into account short- and long-term price fluctuations, making it a useful tool for ensuring reliable bids. This index is essential for owners to estimate the cost of a project, while contractors utilize it to submit financial proposals during the bidding process. The cost of construction is a complex matter that involves frequently changing material and labor costs, as well as other expenses. The volatile economic conditions can potentially cause major disruptions to the company, particularly for large-scale and long-term projects. Hence, careful planning is required to mitigate financial risks and avoid project delays. One of the tools used for predicting future expenses is the CCI [136].

Yaseen and Colleagues propose a novel approach to monthly streamflow prediction utilizing a hybrid evolutionary algorithm that combines adaptive neuro-fuzzy inference systems (ANFIS) and the FFA. Their proposed hybrid ANFIS-FFA model serves as an effective optimizer for the ANFIS model, leading to improved accuracy in streamflow forecasting. To evaluate the forecast accuracy of a new method for monthly streamflow prediction, researchers compare the performance of the proposed hybrid method, which combines an ANFIS with the FFA, to the conventional ANFIS model. Statistical metrics such as correlation coefficient (r), RMSE, and MAE are used to assess the effectiveness of the new model in streamflow forecasting. The results shows that the hybrid ANFIS-FFA model, which uses the FFA as an optimizer, outperforms the traditional ANFIS model in terms of accuracy for monthly stream-flow forecasting. The ANFIS-FFA model resulted in a higher correlation coefficient ($r = 1$) and lower root mean square error (RMSE = 0.984) and mean absolute error (MAE = 0.364) compared to the ANFIS model ($r = 0.998$; RMSE = 3.276; MAE = 1.553). Moreover, the ANFIS-FFA model achieved the improved performance with fewer input variables, demonstrating a more parsimonious approach to stream-flow forecasting. There is

a growing belief that the FFA optimizer can offer better accuracy than the traditional ANFIS model and eliminate any inaccurate data. The hybrid ANFIS-FFA model, which utilizes the FFA algorithm as an optimizer, has shown superior predictive accuracy, indicating that it has far-reaching implications for hydro-meteorological forecasting and other variables that rely solely on historical data inputs. This approach can be used for various types of forecasting beyond streamflow forecasting [137].

Zhou et al., employed the GA and PSO to optimize the input parameters of the ANFIS model for predicting the particle size distribution of muck piles generated from blasting activities. The authors implemented GA and harmony search algorithm (HSA) to optimize the input parameters of ANFIS for predicting the particle size distribution of muck piles after blasting. To validate the accuracy of the proposed ANFIS-GA and ANFIS-HSA models, their performance was compared with that of other regression techniques, including multiple linear regression (MLR), ANN, and DT models. The accuracy of the models was assessed by comparing the predicted fragmentation values with actual fragmentation data obtained from 88 blasting operations conducted in two quarries in Iran. Based on the results, the ANFIS-FFA and ANFIS-GA models performed satisfactorily in predicting fragmentation values, but the ANFIS-GA model outperformed the others with the smallest RMSE and the highest R2 values. The R2 and RMSE values for the ANFIS-GA, ANFIS-FFA, ANFIS, ANN, and SVR models were (0.990 and 0.974), (0.98 and 1.249), (0.95 and 1.591), (0.94 and 2.554), and (0.92 and 2.016), respectively, indicating that the ANFIS-GA model was the most accurate [138].

Madvar et al., undertook a research project aimed at developing an effective approach to predict streamflow forecasting over varying time frames, using a variety of evolutionary algorithms. The authors used several evolutionary algorithms, such as grey wolf optimization (GWO), FFA, PSO, differential evolution (DE), and GA, in conjunction with ANFIS to accurately forecast stream-flow over short- to long-term periods. The findings revealed that all hybrid algorithms significantly outperformed the conventional ANFIS model, with the ANFIS-GWO models demonstrating the best performance. The uncertainty analysis also showed a substantial decrease in uncertainty levels for almost all hybrid models compared to the conventional ANFIS model. Moreover, a significant

advantage of the hybrid ANFIS models developed in the study is the development of a straightforward, explicit equation for streamflow prediction. This is in contrast to traditional Blackbox machine learning models, which do not provide an explicit equation for prediction and require complex calculations to generate forecasts. [139].

Ehteram et al. utilize the ANFIS model to predict infiltration rates, three optimization techniques incorporated with the ANFIS model. ANFIS parameters were tuned using Sine Cosine Algorithm (SCA), PSO and FFA. In this work, experimental results from six separate investigations conducted in various nations were utilized to validate the suggested model. The results suggest that ANFIS-SCA has the potential to provide a more precise evaluation of the infiltration rate compared to ANFIS-PSO. The ANFIS-SCA model showed significantly better results in terms of percent bias errors (PBIAS)and MAE compared to both the ANFIS-FFA and the ANFIS-PSO models, with PBIAS of 0.12 and MAE of 0.007 m3/m. ANIFS-SCA also surpassed ANFIS-FFA in terms of accuracy. In calculating the infiltration rate, the suggested hybrid ANFIS-SCA outperformed previous optimizer algorithms and may be used in various irrigation systems for more sustainable irrigation management [140].

Zhang et al. developed a novel ensemble prediction system that incorporates the decomposition strategy, ensemble point and interval prediction, sub-model selection, and the multi-objective AOA to provide accurate wind speed forecasts. The system was evaluated using hourly-resolution wind speed data from three datasets for the Shandong Peninsula in China. The simulation results indicate that the proposed system outperforms other comparable models in terms of accuracy and reliability, demonstrating its effectiveness for wind speed prediction at both empirical and theoretical levels [141].

Fathy et al. introduced a novel approach based on the AOA to model the maximum power point tracker (MPPT) of wind energy systems. The performance of the proposed AOA-MPPT model was compared with that of other optimization algorithms, including grasshopper optimization algorithm, cuckoo search (CS), and electric charged particle optimization. The simulation results indicated that the

AOA-MPPT model outperformed the other algorithms in terms of optimizing the performance of the wind energy system [120].

Li et al. conducted a study to address the impact of distributed generation (DG) links on power loss and voltage profile, and proposed a new optimization technique called the Integrated Archimedes Optimization Algorithm (IAOA). The IAOA simultaneously determines the locations and capacities of DG units using integer and real values and incorporates an integer mesh-network coding approach to solve the network reconfiguration problem. The effectiveness of the proposed method was compared with PSO and the improved genetic algorithm (IGA), and the results demonstrated that IAOA outperforms the other methods in improving the voltage profile and reducing power loss [61].

Sihag et al. conducted a study on estimating the K value of soil in the Ghaggar basin area using two hybrid approaches, namely the ANN-based ANFIS was combined with particle swarm optimization (ANFIS-PSO) and firefly algorithms (ANFIS-FFA). Field tests were conducted using a minidisk infiltrometer, and 240 observations were recorded, which were divided into two halves. One half was used to construct models, while the other half was used to evaluate the performance of the generated models using four well-known statistical parameters. The results indicate that both ANFIS-FFA and ANFIS-PSO models showed higher precision in predicting the K value of the soil than the classic ANFIS model, with their performance being equivalent [43].

Mustafa and colleagues utilized various meta-heuristic optimization methods (including GWO, PSO, FFA, and GA) combined with the ANFIS to create a machine learning model for predicting the factor of safety against overturning of a gravity retaining wall. The study focused on developing four hybrid models (ANFIS-GWO, ANFIS-PSO, ANFIS-GA, and ANFIS-FFA) and evaluated their performance using different input combinations (cohesiveness, angle of shearing resistance, and unit weight of soil) on two different datasets (70% for training and 30% for testing). Results showed that ANFIS-PSO outperformed the other models in terms of overall performance [142].

4
RESULTS AND DISCUSSION

This section explores the application of hybrid predictive models, including ANFIS, ANN, FFA-ANFIS, AOA-ANFIS, FFA-ANN, and AOA-ANN, for simulating the EAC in construction projects. A critical step in the predictive process involved collecting the cost data from the selected projects by calculating monthly discrepancies between planned and actual costs. The intelligent models were applied to establish mathematical relationships between ten specific attributes—representing various input combinations—and the target variable, EAC. The adoption of these hybrid models within this computational framework mitigates the limitations inherent in traditional indexed formulations. Notably, hybrid intelligence models have proven effective in enhancing human cognitive abilities in addressing complex, real-world problems. Various statistical indicators were employed to quantify absolute error and goodness of fit, as presented in **Tables 4.1-4.9**.

4.1 Traditional Predictive Models

The primary objective of the current research is the proposition of the hybrid predictive models as data intelligence models for the investigated cost at completion problem and to investigate these objective traditional models is obtained for comparison purpose. The indicators of the prediction performances of the traditional ANFIS and ANN-based model using all the input variables are presented in **Table 4.1.**

The analysis of the numerical indicators presented in **Table 4.1** demonstrates that the ANN significantly outperforms the ANFIS model with respect to accuracy. Specifically, the ANN exhibited RMSE and MAE values of 0.131 and 0.092, respectively, alongside an R value of 0.91. In contrast, the ANFIS model recorded RMSE and MAE values of 0.155 and 0.105, with an R value of 0.88. These findings

indicate that the ANN model provides superior performance compared to the ANFIS model. **Figure 4.1** demonstrates (scatter plot) for the relation between predicted EAC and Actual EAC. The first graph refers to using ANN model to represent the Variables, R^2 value was 0.8349 and line formula (y = 0.8476x + 0.0405), while **Figure 4.2** refer to using ANFIS model to represent the Variables, R^2 = 0.7818 and line formula (y = 0.7089x + 0.1259), where the determination coefficient (R^2) has a value ranging from 0 to 1. R^2 values near to one presented that data variation explained clearly by the input data. In contrast, far values from one and near to zero indicate that the data variation is not explained clearly by the input data.

Table 4.1 The numerical performance metrics for the ANFIS and ANN predictive models, referred to as the "Base-model versions," during the testing phase of modeling.

Model	MAE	MSE	R square	RMSE	SMAPE	MRE
ANN	0.092787	0.017412	0.8349	0.131954	0.315866	0.469087
ANFIS	0.105993	0.024037	0.7818	0.155040	0.335471	0.418710

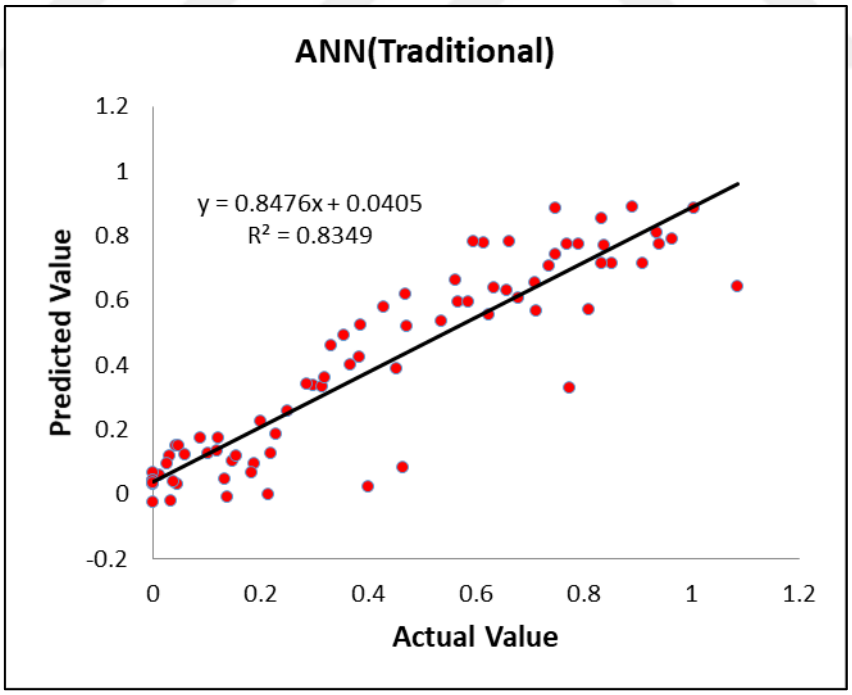


Figure 4.1 The scatter plot diagram imagining between the actual and predicted values of EAC for traditional ANN model

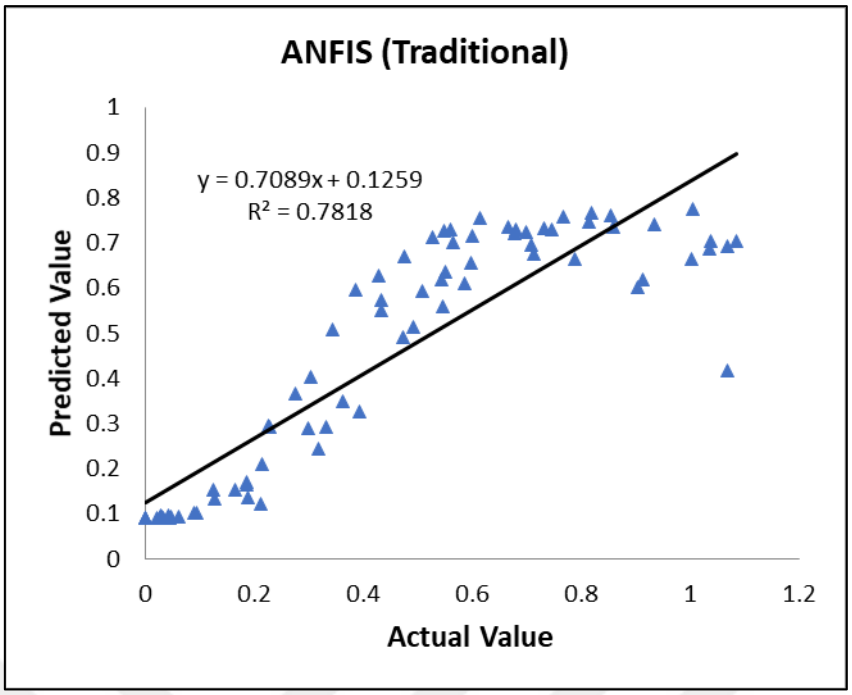


Figure 4.2 The scatter plot diagram imagining between the actual and predicted values of EAC for traditional ANFIS model.

4.2 The Performance of Hybrid Predictive Models

The enthusiasm for coupling the input selection approach to the predictive model is to explore the predominant input combination correlated to the EAC magnitude. Note that, this is highly magnificent to recognize the main influenced variables during the project progress that affect the variance of the EAC results. The FFA &AOA hybridized with the ANN&AFNIS to abstract the suitable input combination. The input combination and the prediction skill results of the hybrid model FFA-ANN are indicated in **Tables 4.2** and **4.3**, respectively.

By studying the archived modeling results in **Table 4.3**, Model 7 exhibited the best input combination for predicting EAC through including (CEI, Progress (%), COI, ACp, SBI, CPI, CCI) variables as inputs for the prediction matrix. The results showed minimum absolute error metrics (e.g. RMSE-MAE) (0.068657- 0.053076) and best-fit-goodness (e.g., R) (0.9766). As shown in **Table 4.3** model 3 also achieved excellent results by using only 3 variables which are (COI, EVp, ACp). The hybrid FFA-ANFIS model behavior illustrated in (**Tables 4.4 and 4.5**), six input variables represented in (CPI, EVp, COI, OBI, ACp, Progress (%)), gave the optimal prediction skills with minimum RMSE = 0.096226, MAE = 0.061982 and

R = 0.955. Note that FF-ANN surpassed the capability of the FFA-ANFIS model; however, with more features for comprehending the internal mapping relationship between predictors and predicted.

Table 4.2 The feature combination elements to predict EAC value using FFA-ANN model

No. of Variables	Models	Input Variables	Output
1 Variable	M- 1	[Progress (%)]	EAC
2 Variables	M- 2	[CCI, OBI]	
3 Variables	M- 3	[COI, EVp, ACp]	
4 Variables	M- 4	[CPI, COI, SPI, OBI]	
5 Variables	M- 5	[CCI, COI, CEI, SPI, ACp]	
6 Variables	M- 6	[SPI, COI, OBI, CEI, Progress (%), EVp]	
7 Variables	**M- 7**	**[CEI, Progress (%), COI, ACp, SBI, CPI, CCI]**	
8 Variables	M- 8	[CEI, SPI, Progress (%), CCI, SBI, OBI, EVp, CPI]	
9 Variables	M- 9	[COI, CPI, SBI, SPI, EVp, CEI, CCI, Progress (%), OBI]	

Table 4.3 The FFA-ANN prediction model's numerical indicators for the deferent combination modeling phase

Model	MAE	MSE	R Square	RMSE	SMAPE	MRE
M-1	0.076264	0.011089	0.8988	0.105305	0.252074	0.291046
M-2	0.219545	0.072027	0.3069	0.268379	0.642433	1.698822
M-3	0.056839	0.005515	0.9505	0.074265	0.177772	0.316342
M-4	0.261234	0.092195	0.099	0.303636	0.726830	2.068755
M-5	0.065280	0.006471	0.9372	0.080445	0.192565	0.374820
M-6	0.063154	0.007761	0.9256	0.088095	0.210877	0.243749
M-7	**0.053076**	**0.004714**	**0.9541**	**0.068657**	**0.164348**	**0.210831**
M-8	0.075951	0.010557	0.8994	0.102747	0.245952	0.371473
M-9	0.054667	0.005654	0.948	0.075192	0.179991	0.216728

Table 4.4 The feature combination elements used to predict EAC value using FFA-ANFIS model

No. of Variables	Models	Input Variables	Output
1 Variable	M- 1	[Progress (%)]	EAC
2 Variables	M- 2	[EVp, Progress (%)]	
3 Variables	M- 3	[CEI, OBI, Progress (%)]	
4 Variables	M- 4	[ACp, OBI, CEI, SPI]	
5 Variables	M- 5	[CCI, EVp, OBI, Progress (%), CPI]	
6 Variables	**M- 6**	**[CPI, EVp, COI, OBI, ACp, Progress (%)]**	
7 Variables	M- 7	[Progress (%), SPI, EVp, SBI, ACp, CPI, CCI]	
8 Variables	M- 8	[Progress (%), CCI, COI, OBI, ACp, SPI, ACp, EVp]	
9 Variables	M- 9	[SBI, ACp, SPI, OBI, CEI, CPI, EVp, Progress (%), COI]	

Table 4.5 The FFA-ANFIS prediction model's numerical indicators for the deferent combination modeling phase

Model	MAE	MSE	R square	RMSE	SMAPE	MRE
M-1	0.099710	0.020721	0.8384	0.143948	0.311471	0.402396
M-2	0.083227	0.016540	0.8528	0.128606	0.278275	0.370809
M-3	0.088432	0.016589	0.8487	0.128799	0.278692	0.281587
M-4	0.115234	0.026565	0.7676	0.162987	0.352667	0.838178
M-5	0.097129	0.019521	0.8262	0.139716	0.302314	0.568085
M-6	**0.061982**	**0.009259**	**0.9159**	**0.096226**	**0.208211**	**0.197332**
M-7	0.142641	0.030103	0.7405	0.173502	0.375419	0.977923
M-8	0.085524	0.016142	0.8561	0.127052	0.274912	0.240471
M-9	0.095588	0.016145	0.8644	0.127063	0.274935	0.440669

Based on the reported prediction accuracy in **Table 4.5**, the performance of the FFA-ANN model was demonstrated a superior capacity to the FFA-ANFIS model. The results of the input combinations variables and the prediction performances of the AOA-ANFIS are displayed in **Tables 4.6 and 4.7**, respectively.

Table 4.6 The feature combination elements used to predict EAC value using AOA-ANFIS model

No. of Variables	Models	Input Variables	Output
1 Variable	M- 1	[CPI]	**EAC**
2 Variables	M- 2	[CPI, EVp]	
3 Variables	M- 3	[Progress (%), SBI, ACp]	
4 Variables	M- 4	[Progress (%), OBI, SBI, SPI]	
5 Variables	M- 5	[Progress (%), SPI, COI, OBI, CEI]	
6 Variables	**M- 6**	**[Progress (%), COI, SPI, ACp, CPI, CCI]**	
7 Variables	M- 7	[Progress (%), CCI, ACp, SPI, EVp, SBI, CPI]	
8 Variables	M- 8	[Progress (%), CCI, EVp, OBI, COI, CPI, SPI, SBI]	
9 Variables	M- 9	[Progress (%), SPI, CPI, OBI, CCI, EVp, ACp, COI, SBI]	

Table 4.7 The AOA-ANFIS prediction model's numerical indicators for the deferent combination modeling phase

Model	MAE	MSE	R square	RMSE	SMAPE	MRE
M-1	0.234069	0.076944	0.166	0.277387	0.636038	1.782503
M-2	0.082314	0.015279	0.8276	0.123607	0.283427	0.397067
M-3	0.084676	0.011597	0.8688	0.107688	0.246924	0.305348
M-4	0.076448	0.010318	0.8836	0.101578	0.232914	0.262587
M-5	0.110348	0.016610	0.8306	0.128878	0.295514	0.781438
M-6	**0.061145**	**0.006390**	**0.9283**	**0.079941**	**0.183301**	**0.289979**
M-7	0.157015	0.042866	0.5358	0.207042	0.474739	0.851820
M-8	0.093761	0.016150	0.8177	0.127082	0.291393	0.437062
M-9	0.081021	0.010281	0.8848	0.101397	0.232500	0.365132

The optimum input combination of the AOA-ANFIS model was performed using the six combinations by including the Progress (%), COI, SPI, ACp, CPI and CCI variables. On the other hand, AOA-ANN allocates its best predictability using the seven-input combination by incorporating the Progress (%), OBI, SBI, SPI, CPI, COI and EVp variables (**Tables 4.8 and 4.9**). For more convenience, a comparative analysis was performed between the FFA-ANFIS and FFA-ANN models and the AOA-ANFIS and AOA-ANN models. The comparison showed that the FFA-ANN

model was superior in terms of significant improvements based on the quantitative measurements. However, it should be noticed that the AOA related models required more execution time to abstract the internal relationship between the predictors and the predicted.

Table 4.8 The feature combination elements used to predict EAC value using AOA-ANN model

No. of Variables	Models	Input Variables	Output
1 Variable	M- 1	[Progress (%)]	EAC
2 Variables	M- 2	[CPI, EVp]	
3 Variables	M- 3	[Progress (%), COI, CPI]	
4 Variables	M- 4	[CPI, OBI, SPI, EVp]	
5 Variables	M- 5	[Progress (%), OBI, CPI, SPI, SBI]	
6 Variables	M- 6	[Progress (%), ACp, CCI, EVp, CPI, OBI]	
7 Variables	**M- 7**	**[Progress (%), OBI, SBI, SPI, CPI, COI, EVp]**	
8 Variables	M- 8	[Progress (%), OBI, SPI, EVp, ACp, CPI, COI, SBI]	
9 Variables	M- 9	[Progress (%), EVp, CCI, SPI, COI, ACp, CPI, OBI, SBI]	

Table 4.9 The AOA-ANN prediction model's numerical indicators for the deferent combination modeling phase

Model	MAE	MSE	R square	RMSE	SMAPE	MRE
M-1	0.080080	0.010614	0.9025	0.103026	0.238982	0.467914
M-2	0.097747	0.014573	0.872	0.120717	0.280019	1.142387
M-3	0.087326	0.011451	0.8703	0.107012	0.248227	0.517756
M-4	0.091338	0.014366	0.8703	0.119857	0.278024	1.020044
M-5	0.288813	0.108470	0.0254	0.329348	0.763965	3.316843
M-6	0.073331	0.009339	0.9209	0.096637	0.224162	0.308758
M-7	**0.056581**	**0.005828**	**0.9465**	**0.076341**	**0.177082**	**0.270783**
M-8	0.063507	0.006942	0.9401	0.083318	0.193267	0.378148
M-9	0.074529	0.009164	0.9351	0.095727	0.222050	0.489055

Another way of evaluation usually used to visualize the predictive models' capability is a scatter plot. The scatter plot or the variation from the best fit line is

a graphical way of representing the relationship between actual and predicted values. **Figures 4.3-4.2** showed the deviation from the ideal 45° line for the hybrid AFNIS and ANN models.

In **Figure 4.3**, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 1. This comparison involves variable (Progress (%)) for both FFA-ANFIS and FFA-ANN models Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. FFA-ANN yielded an R2 value of 0.8988, while FFA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.8384. This stark contrast emphasizes the superior determination coefficient (R^2) achieved by FFA-ANN compared to FFA-ANFIS when using the same set of variables.

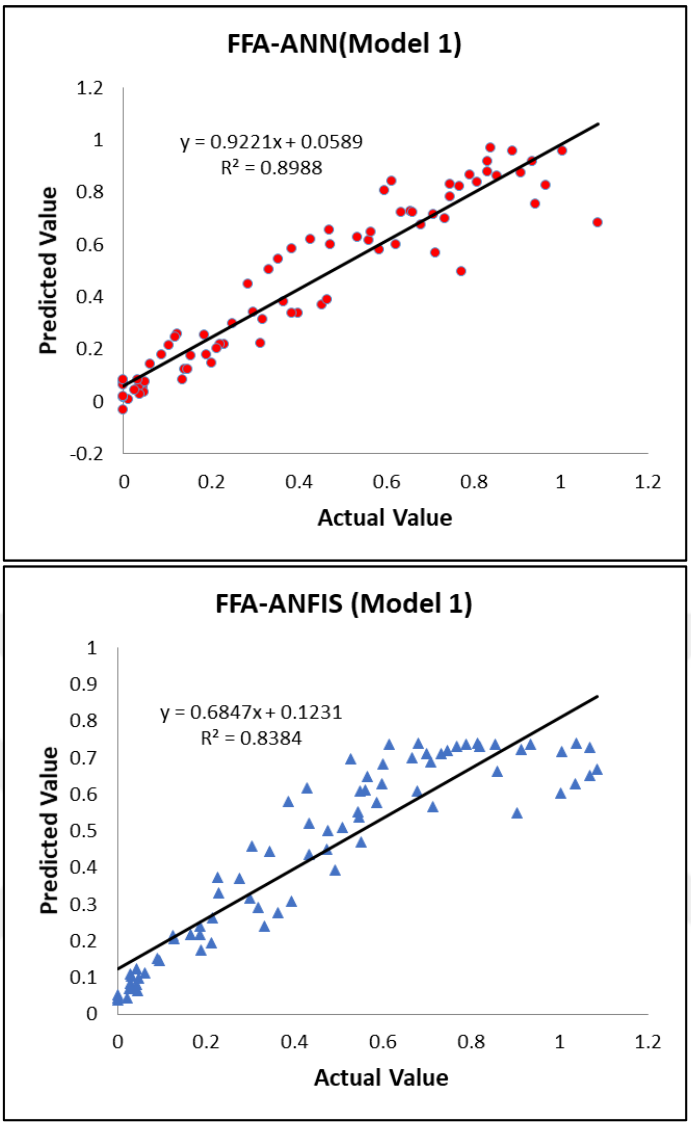


Figure 4.3 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 1 (FFA-ANN &FF-ANFIS).

In **Figure 4.4**, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 2. This comparison involves variables (CCI, OBI) for FFA-ANN model and (EVp, Progress (%)) for FFA-ANFIS model. Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. FFA-ANN yielded an R^2 value of 0.306, while FFA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.852. This stark contrast emphasizes the superior determination coefficient (R^2) achieved by FFA-ANFIS compared to FFA-ANN when using the same set of variables.

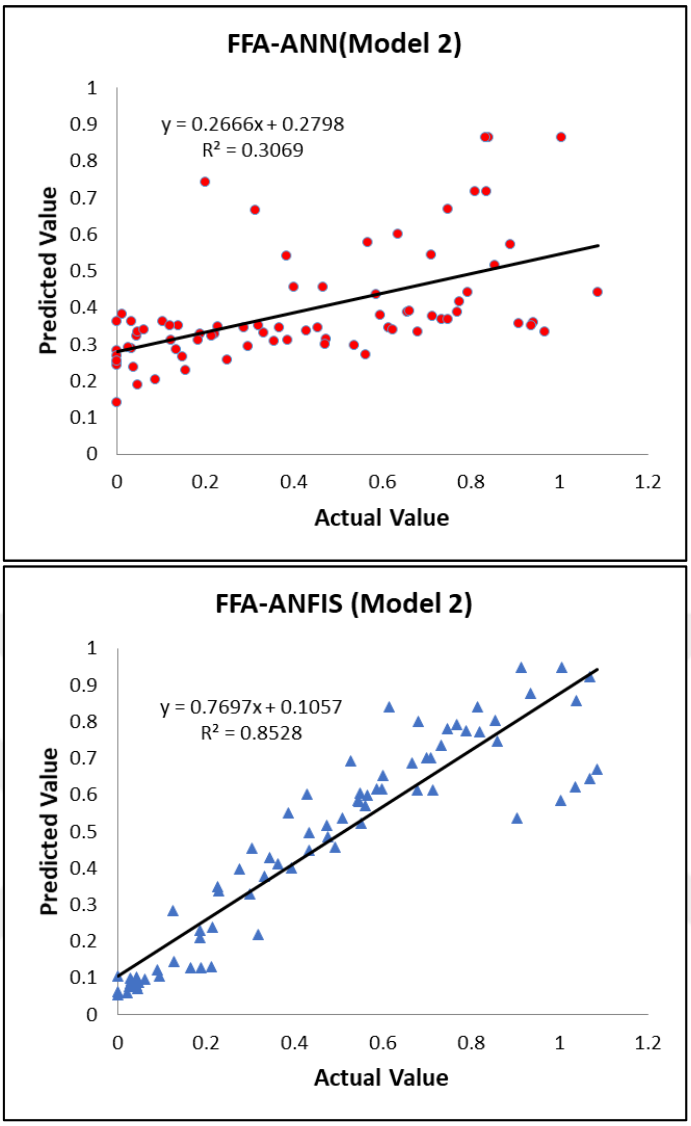


Figure 4.4 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 2 (FFA-ANN &FF-ANFIS).

In **Figure 4.5**, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 3. This comparison involves variables (COI, EVp, ACp) for FFA-ANN model and (CEI, OBI, Progress (%)) for FFA-ANFIS model. Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. FFA-ANN yielded an R^2 value of 0.95, while FFA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.848. This stark contrast emphasizes the superior determination coefficient (R^2) achieved by FFA-ANN compared to FFA-ANFIS when using the same set of variables.

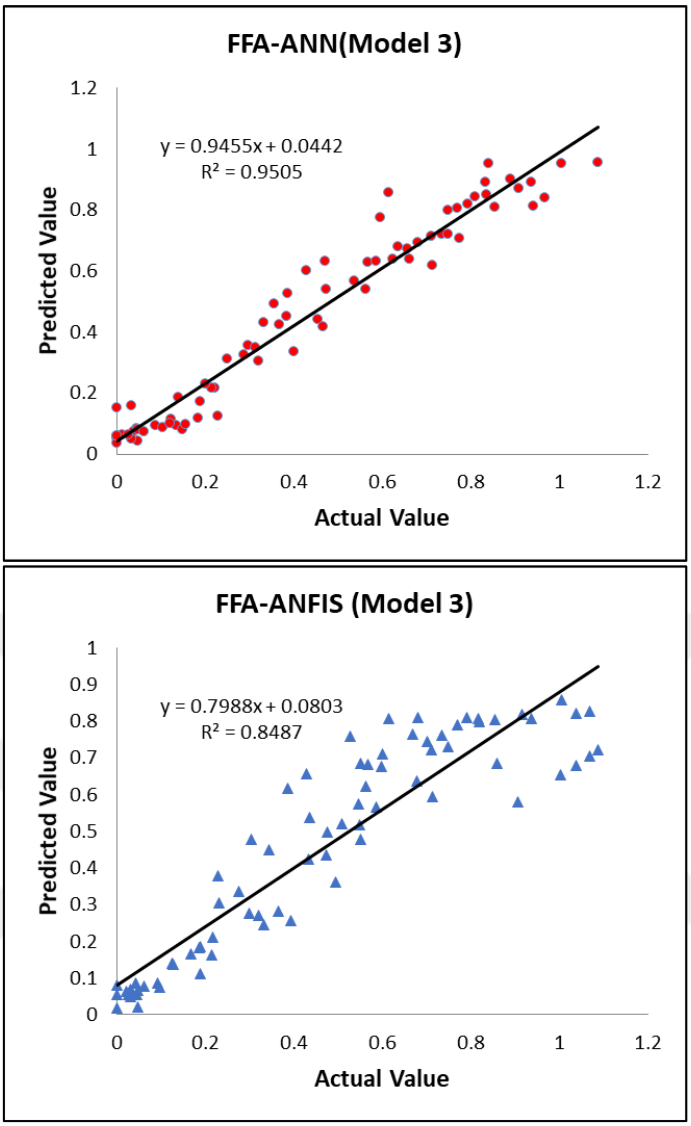


Figure 4.5 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 3 (FFA-ANN &FF-ANFIS).

In **Figure 4.6**, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 4. This comparison involves variables (CPI, COI, SPI, OBI) for FFA-ANN model and (ACp, OBI, CEI, SPI) for FFA-ANFIS model. Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. FFA-ANN yielded an R^2 value of 0.099, while FFA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.767. This stark contrast emphasizes the superior determination coefficient (R^2) achieved by FFA-ANFIS compared to FFA-ANN when using the same set of variables.

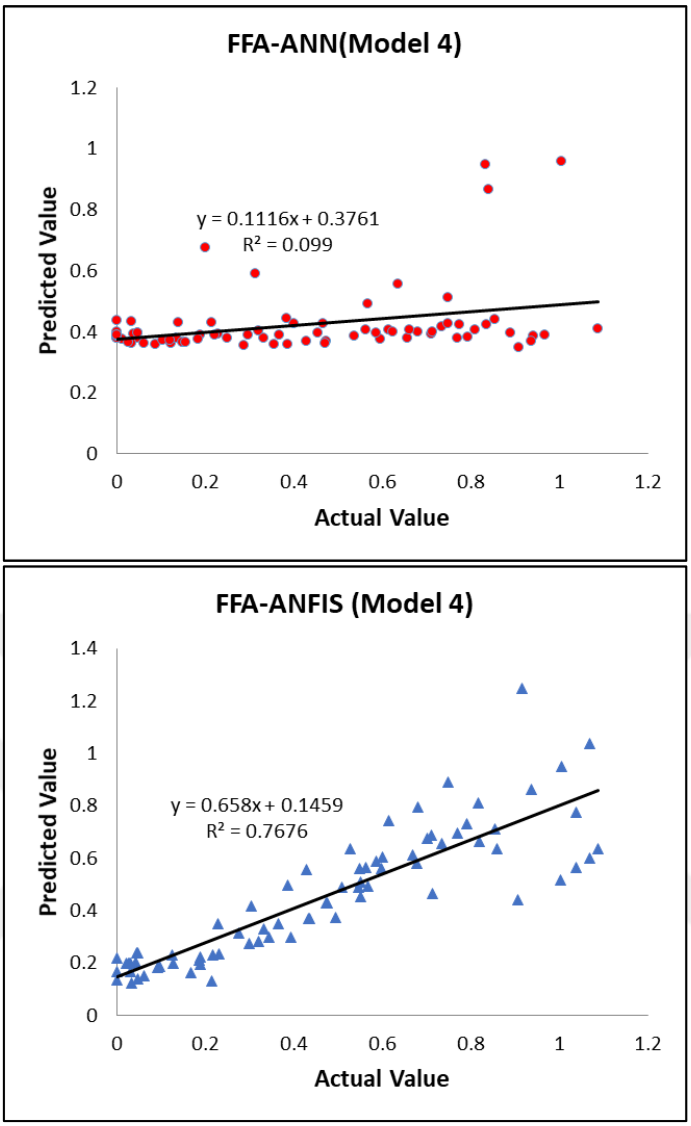


Figure 4.6 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 4 (FFA-ANN &FF-ANFIS).

In **Figure 4.7**, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 5. This comparison involves variables (CCI, COI, CEI, SPI, ACp) for FFA-ANN model and (CCI, EVp, OBI, Progress (%), CPI) for FFA-ANFIS model. Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. FFA-ANN yielded an R^2 value of 0.937, while FFA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.826. This stark contrast emphasizes the superior determination coefficient (R^2) achieved by FFA-ANN compared to FFA-ANFIS when using the same set of variables.

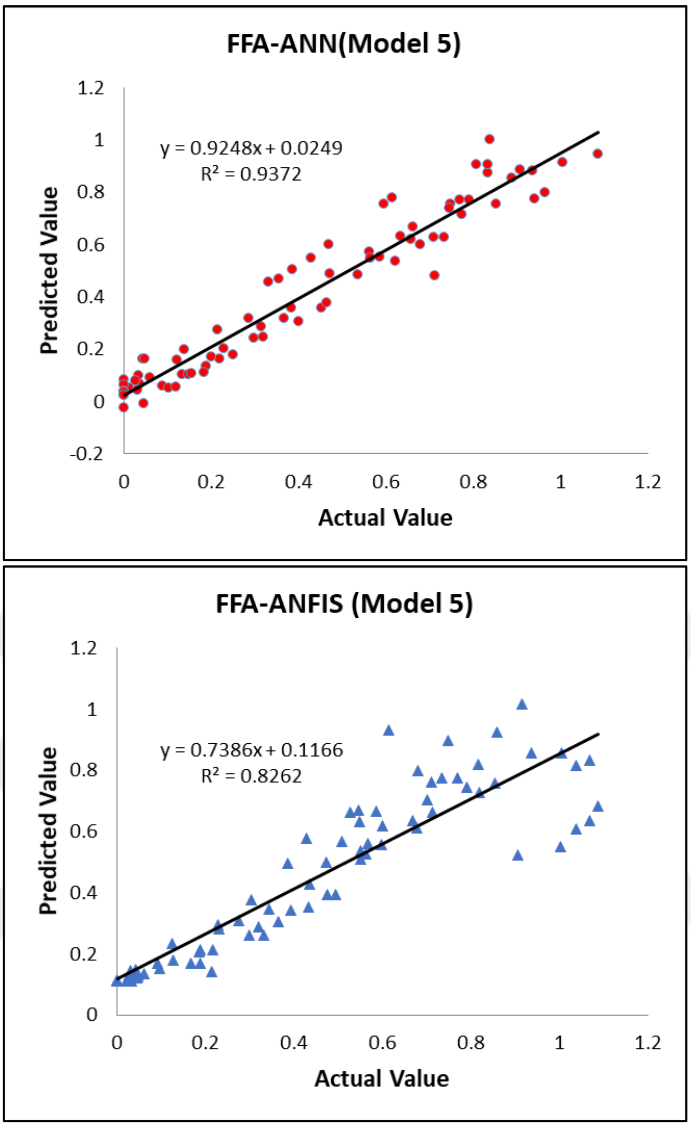


Figure 4.7 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 5 (FFA-ANN &FF-ANFIS).

In **Figure 4.8**, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 6. This comparison involves variables (SPI, COI, OBI, CEI, Progress (%), EVp) for FFA-ANN model and (CPI, EVp, COI, OBI, ACp, Progress (%)) for FFA-ANFIS model. Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. FFA-ANN yielded an R^2 value of 0.925, while FFA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.915. This stark contrast emphasizes the superior determination coefficient (R^2) achieved by FFA-ANN compared to FFA-ANFIS when using the same set of variables.

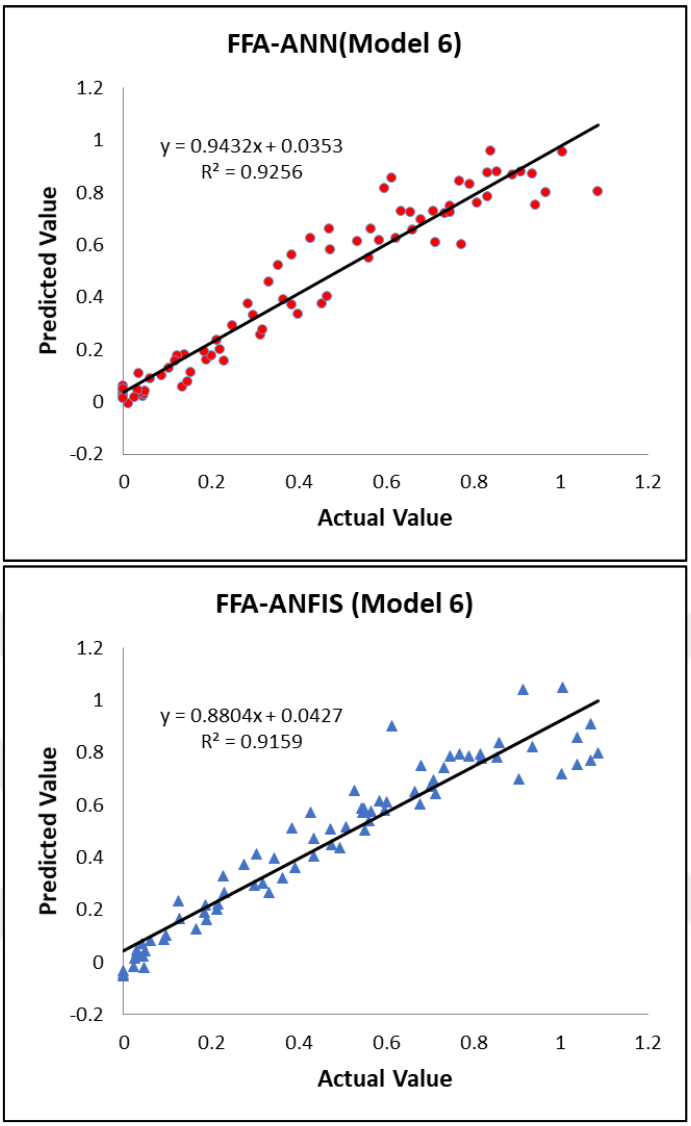


Figure 4.8 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 6 (FFA-ANN &FF-ANFIS).

In **Figure 4.9**, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 7. This comparison involves variables (CEI, Progress (%), COI, ACp, SBI, CPI, CCI) for FFA-ANN model and (Progress (%), SPI, EVp, SBI, ACp, CPI, CCI) for FFA-ANFIS model. Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. FFA-ANN yielded an R^2 value of 0.954, while FFA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.740. This stark contrast emphasizes the superior determination coefficient (R^2) achieved by FFA-ANN compared to FFA-ANFIS when using the same set of variables.

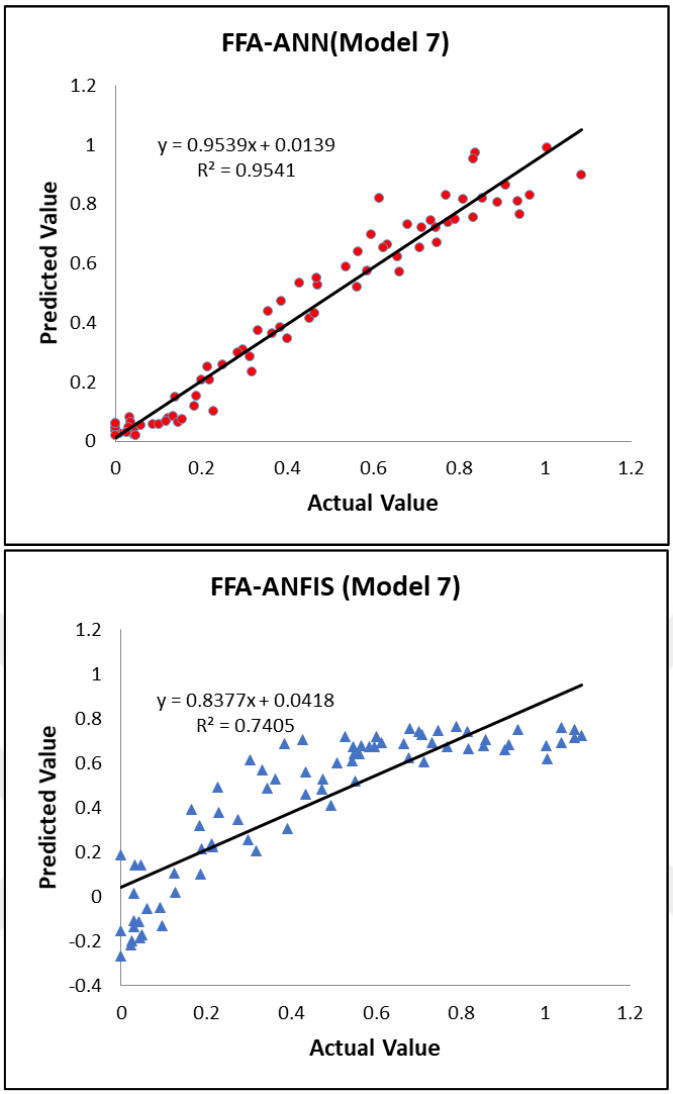


Figure 4.9 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 7 (FFA-ANN &FF-ANFIS).

In **Figure 4.10**, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 8. This comparison involves variables (CEI, SPI, Progress (%), CCI, SBI, OBI, EVp, CPI) for FFA-ANN model and (Progress (%), CCI, COI, OBI, ACp, SPI, ACp, EVp) for FFA-ANFIS model. Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. FFA-ANN yielded an R^2 value of 0.899, while FFA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.856. This stark contrast emphasizes the superior determination coefficient (R^2) achieved by FFA-ANN compared to FFA-ANFIS when using the same set of variables.

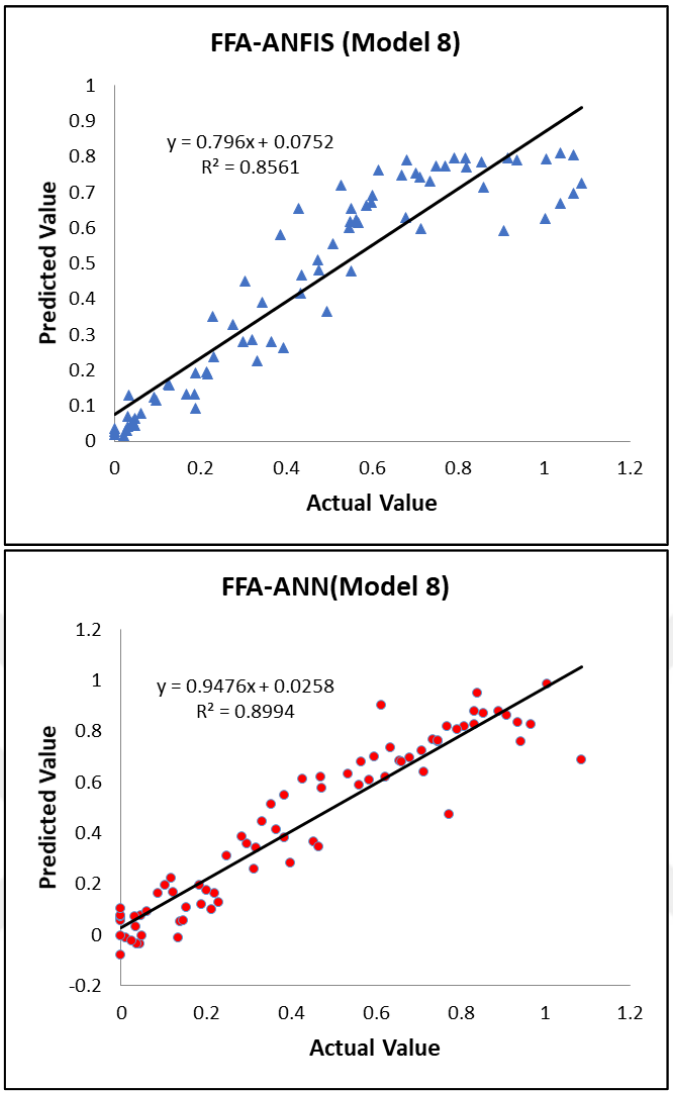


Figure 4.10 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 8 (FFA-ANN &FF-ANFIS).

In **Figure 4.11**, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 9. This comparison involves variables (COI, CPI, SBI, SPI, EVp, CEI, CCI, Progress (%), OBI) for FFA-ANN model and (COI, CPI, SBI, SPI, EVp, CEI, CCI, Progress (%), OBI) for FFA-ANFIS model. Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. FFA-ANN yielded an R^2 value of 0.948, while FFA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.864. This stark contrast emphasizes the superior determination coefficient (R^2) achieved by FFA-ANN compared to FFA-ANFIS when using the same set of variables.

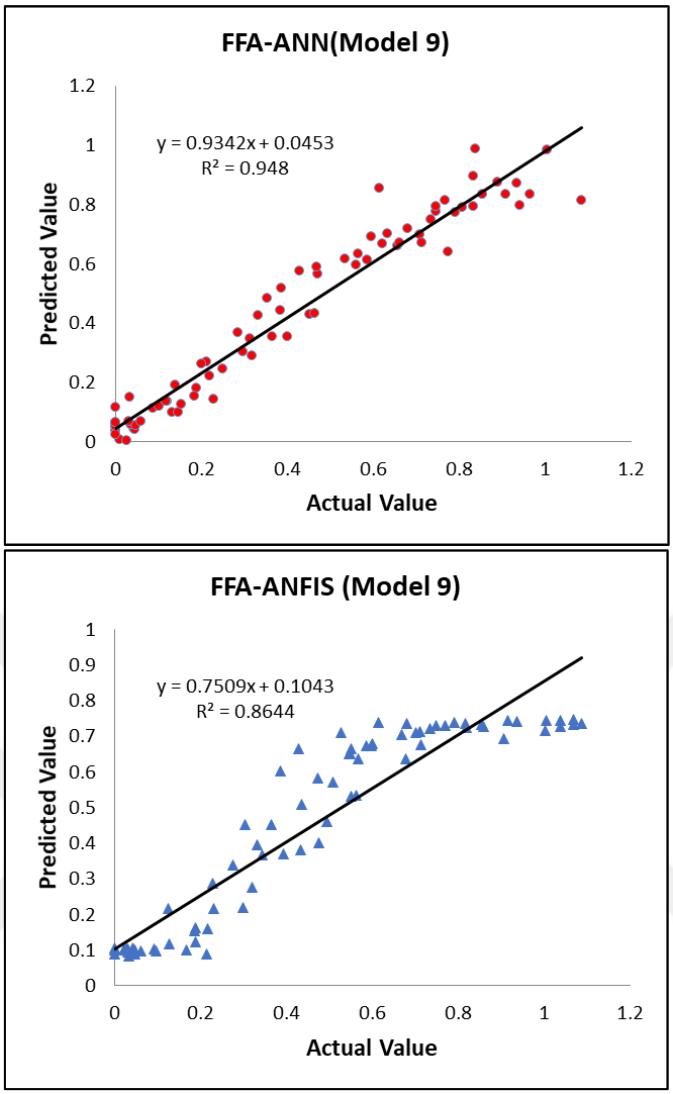


Figure 4.11 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 9 (FFA-ANN &FF-ANFIS).

In **Figure 4.12**, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 1. This comparison involves variable (Progress (%)) for AOA-ANN model and (CPI) for AOA-ANFIS model. Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. AOA-ANN yielded an R^2 value of 0.902, while AOA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.166. This stark contrast emphasizes the superior determination coefficient (R^2) achieved by AOA-ANN compared to AOA-ANFIS when using the same set of variables.

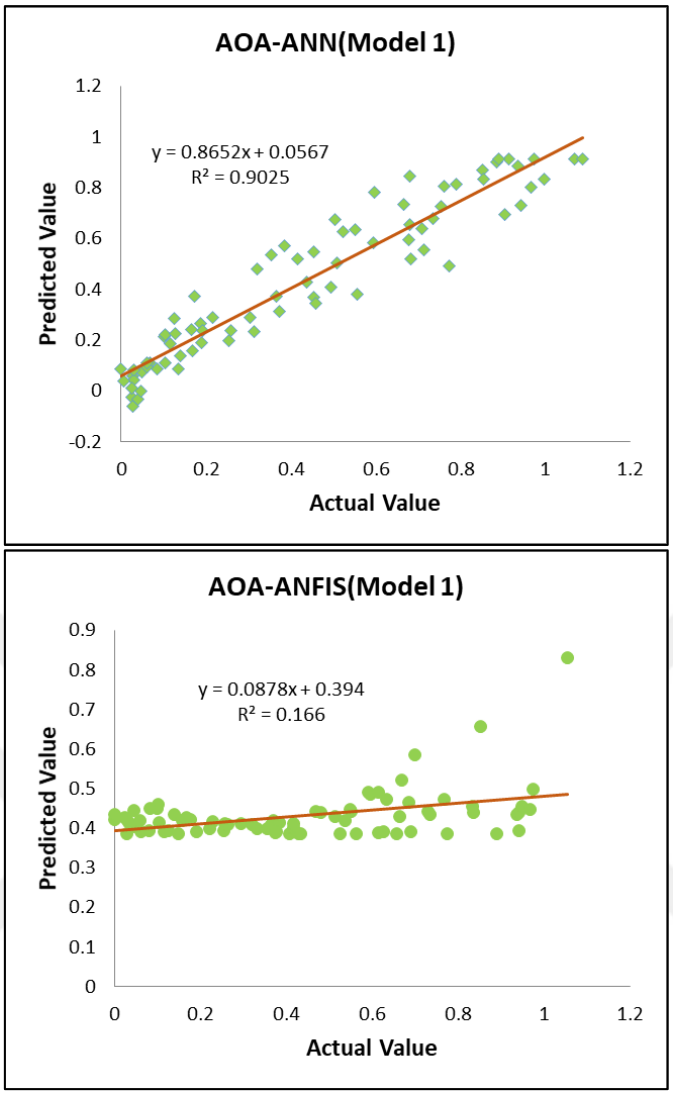


Figure 4.12 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 1 (AOA-ANN &AOA-ANFIS).

In **Figure 4.13**, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 2. This comparison involves variables (CPI, EVp) for both AOA-ANN and AOA-ANFIS models. Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. AOA-ANN yielded an R^2 value of 0.872, while AOA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.827. This stark contrast emphasizes the superior determination coefficient (R^2) achieved by AOA-ANN compared to AOA-ANFIS when using the same set of variables.

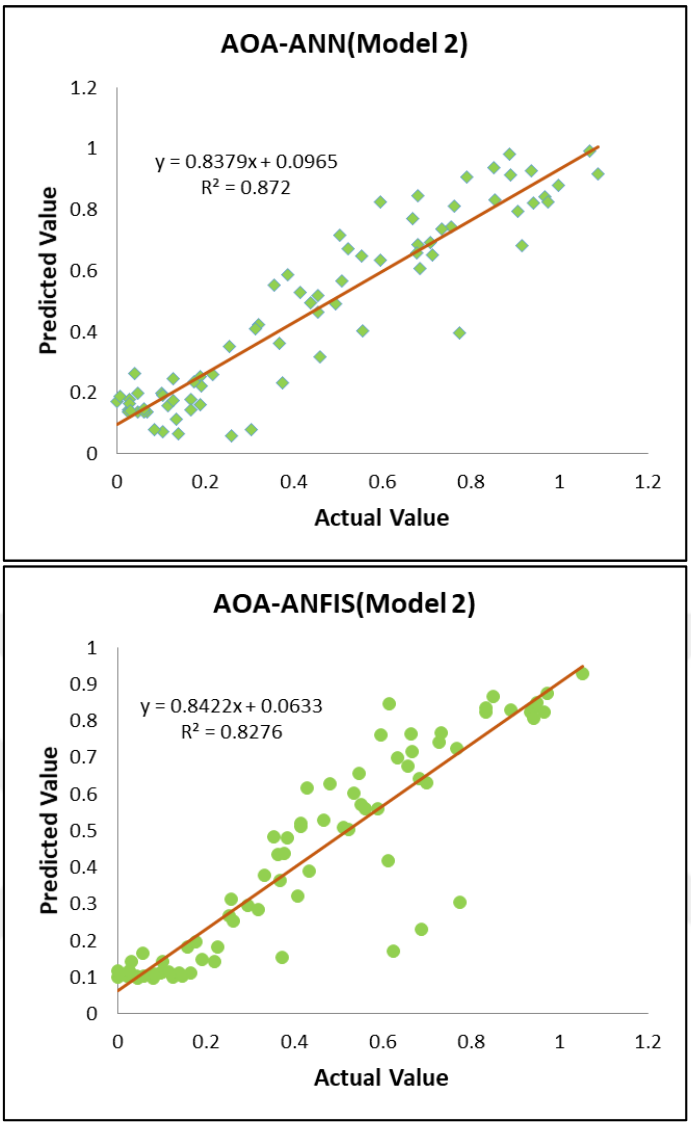


Figure 4.13 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 2 (AOA-ANN &AOA-ANFIS).

In **Figure 4.14**, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 3. This comparison involves variables (Progress (%), COI, CPI) for AOA-ANN model and (Progress (%), SBI, ACp) for AOA-ANFIS model. Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. AOA-ANN yielded an R^2 value of 0.896, while AOA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.868. This stark contrast emphasizes the superior determination coefficient (R^2) achieved by AOA-ANN compared to AOA-ANFIS when using the same set of variables.

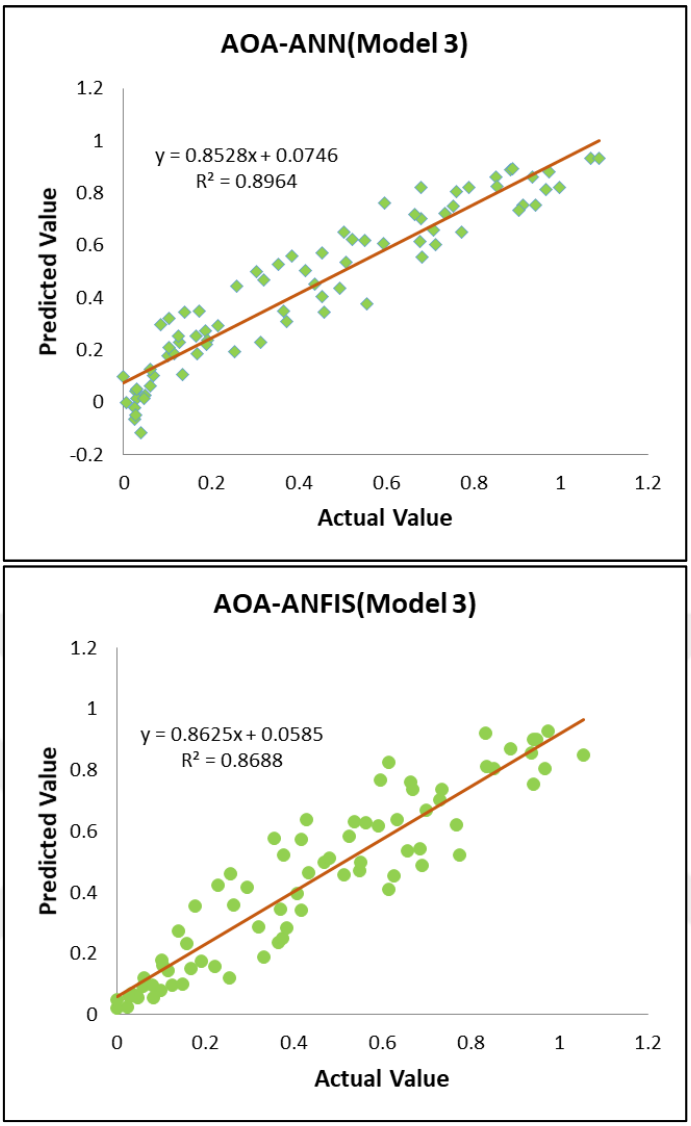


Figure 4.14 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 3 (AOA-ANN &AOA-ANFIS).

In **Figure 4.15**, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 4. This comparison involves variables (CPI, OBI, SPI, EVp) for AOA-ANN model and (Progress (%), OBI, SBI, SPI) for AOA-ANFIS model. Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. AOA-ANN yielded an R^2 value of 0.870, while AOA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.883. This stark contrast emphasizes the superior determination coefficient (R^2) achieved by AOA-ANFIS compared to AOA-ANN when using the same set of variables.

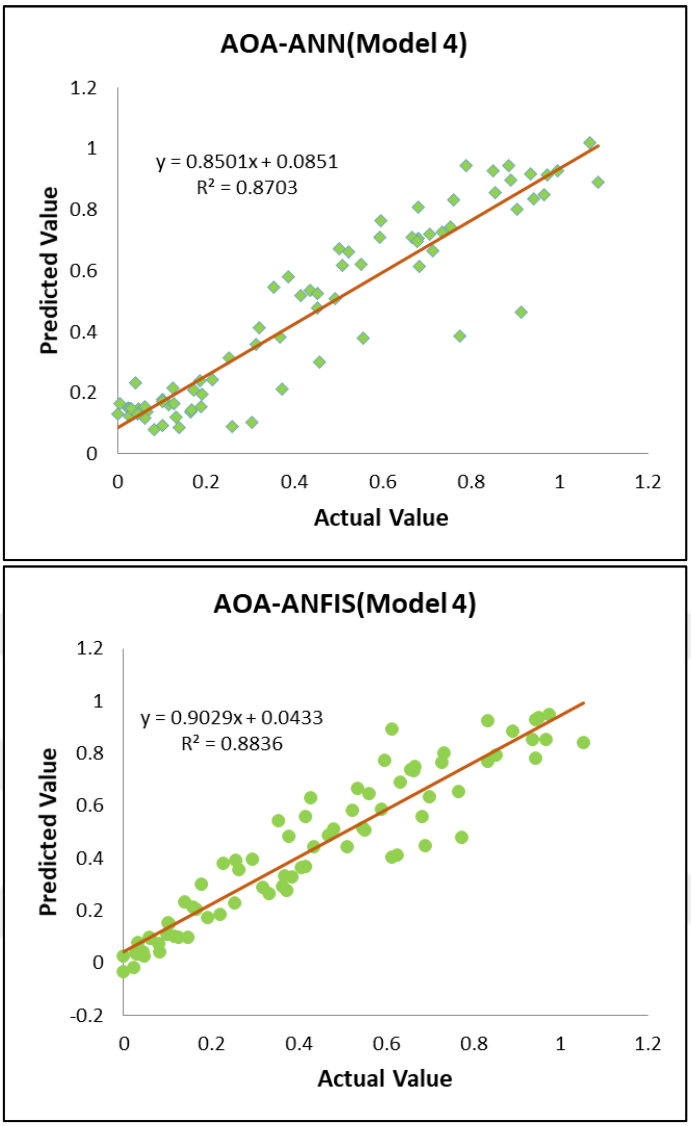


Figure 4.15 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 4 (AOA-ANN &AOA-ANFIS).

In **Figure 4.16**, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 5. This comparison involves variables (Progress (%), OBI, CPI, SPI, SBI) for AOA-ANN model and (Progress (%), SPI, COI, OBI, CEI) for AOA-ANFIS model. Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. AOA-ANN yielded an R^2 value of 0.025, while AOA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.830. This stark contrast emphasizes the superior determination coefficient (R^2) achieved by AOA-ANFIS compared to AOA-ANN when using the same set of variables.

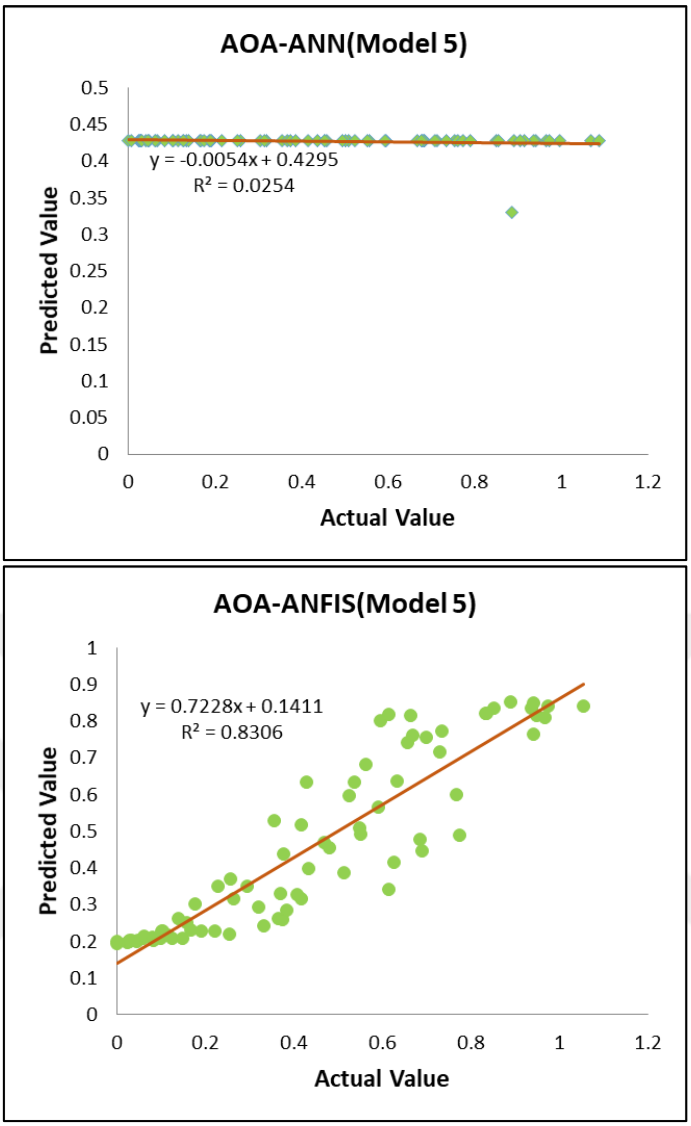


Figure 4.16 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 5 (AOA-ANN &AOA-ANFIS).

In **Figure 4.17**, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 6. This comparison involves variables (Progress (%), ACp, CCI, EVp, CPI, OBI) for AOA-ANN model and (Progress (%), COI, SPI, ACp, CPI, CCI) for AOA-ANFIS model. Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. AOA-ANN yielded an R^2 value of 0.920, while AOA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.928. This stark contrast emphasizes that approximately same determination coefficient (R^2) achieved by both AOA-ANFIS and AOA-ANN when using the same set of variables, although deferent features are used for each model.

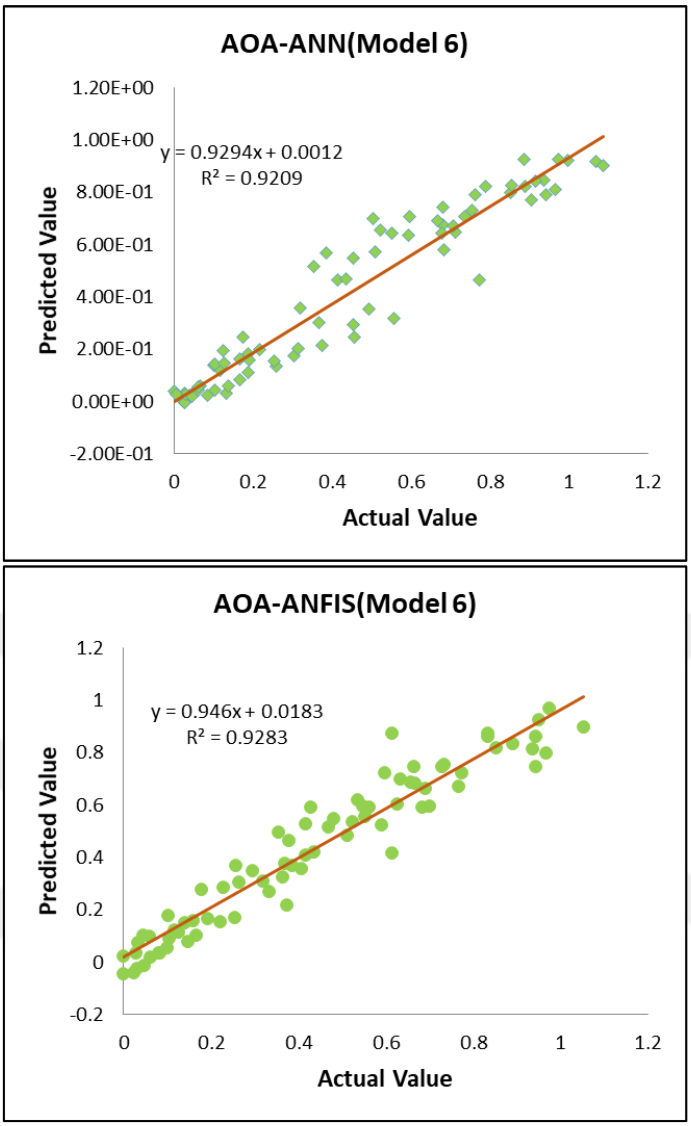


Figure 4.17 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 6 (AOA-ANN &AOA-ANFIS).

In **Figure 4.18**, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 7. This comparison involves variables (Progress (%), OBI, SBI, SPI, CPI, COI, EVp) for AOA-ANN model and (Progress (%), CCI, ACp, SPI, EVp, SBI, CPI) for AOA-ANFIS model. Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. AOA-ANN yielded an R^2 value of 0.946, while AOA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.535. This stark contrast emphasizes the superior determination coefficient (R^2) achieved by AOA-ANN compared to AOA-ANFIS when using the same set of variables.

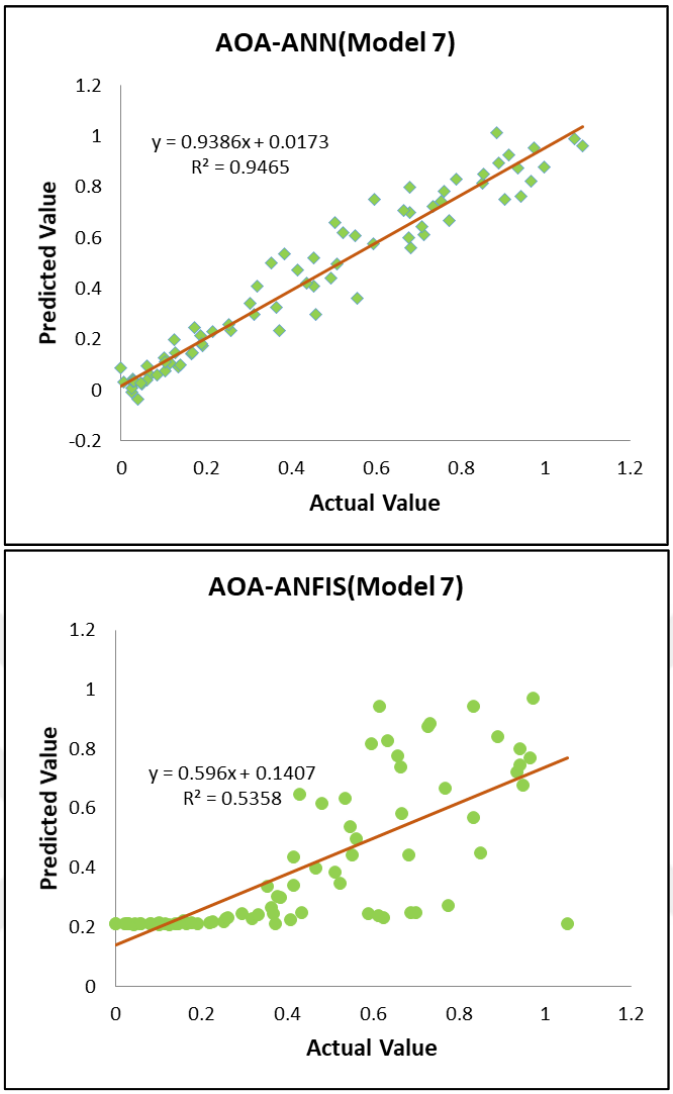


Figure 4.18 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 7 (AOA-ANN &AOA-ANFIS).

In **Figure 4.19**, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 8. This comparison involves variables (Progress (%), OBI, SPI, EVp, ACp, CPI, COI, SBI) for AOA-ANN model and (Progress (%), CCI, EVp, OBI, COI, CPI, SPI, SBI) for AOA-ANFIS model. Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. AOA-ANN yielded an R^2 value of 0.940, while AOA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.817. This stark contrast emphasizes the superior determination coefficient (R^2) achieved by AOA-ANN compared to AOA-ANFIS when using the same set of variables.

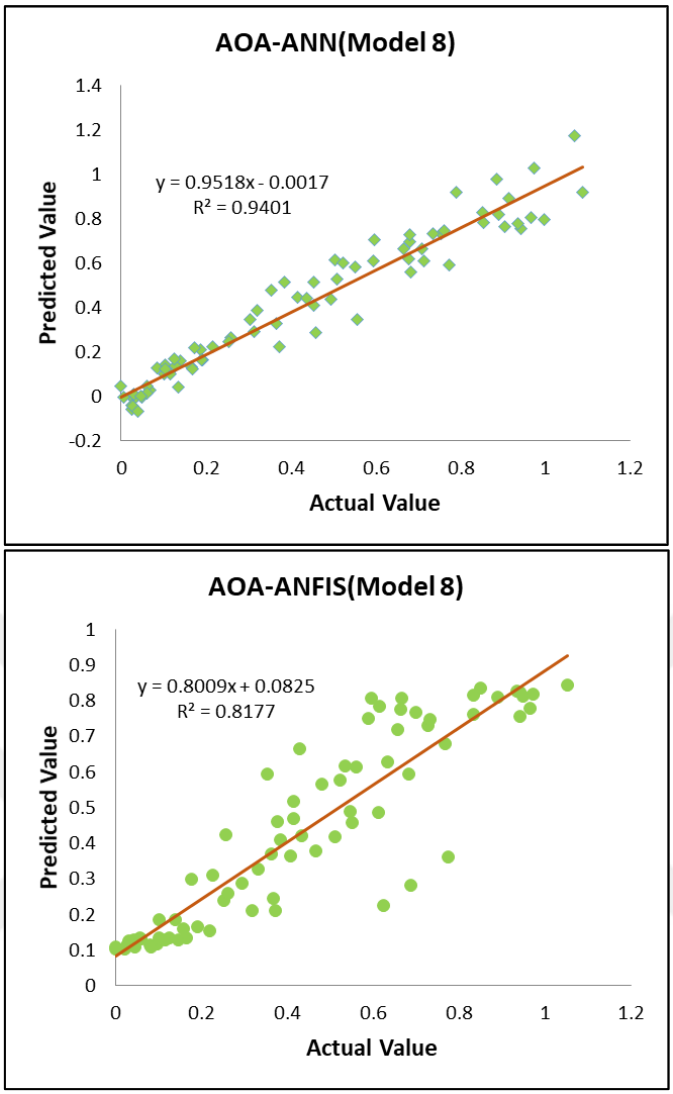


Figure 4.19 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 8 (AOA-ANN &AOA-ANFIS).

In Figure 4.20, the relationship between predicted Estimate at Completion (EAC) and actual EAC is illustrated for both FFA-ANFIS and FFA-ANN predictive models in Model 9. This comparison involves variables (Progress (%), EVp, CCI, SPI, COI, ACp, CPI, OBI, SBI) for AOA-ANN model and (Progress (%), SPI, CPI, OBI, CCI, EVp, ACp, COI, SBI) for AOA-ANFIS model. Through multiple attempts employing mathematical equations to optimize the model, the determination coefficient (R^2) value and the line formula were derived. AOA-ANN yielded an R^2 value of 0.935, while AOA-ANFIS, representing the variable, resulted in a lower R^2 value of 0.884. This stark contrast emphasizes the superior determination coefficient (R^2) achieved by AOA-ANN compared to AOA-ANFIS when using the same set of variables.

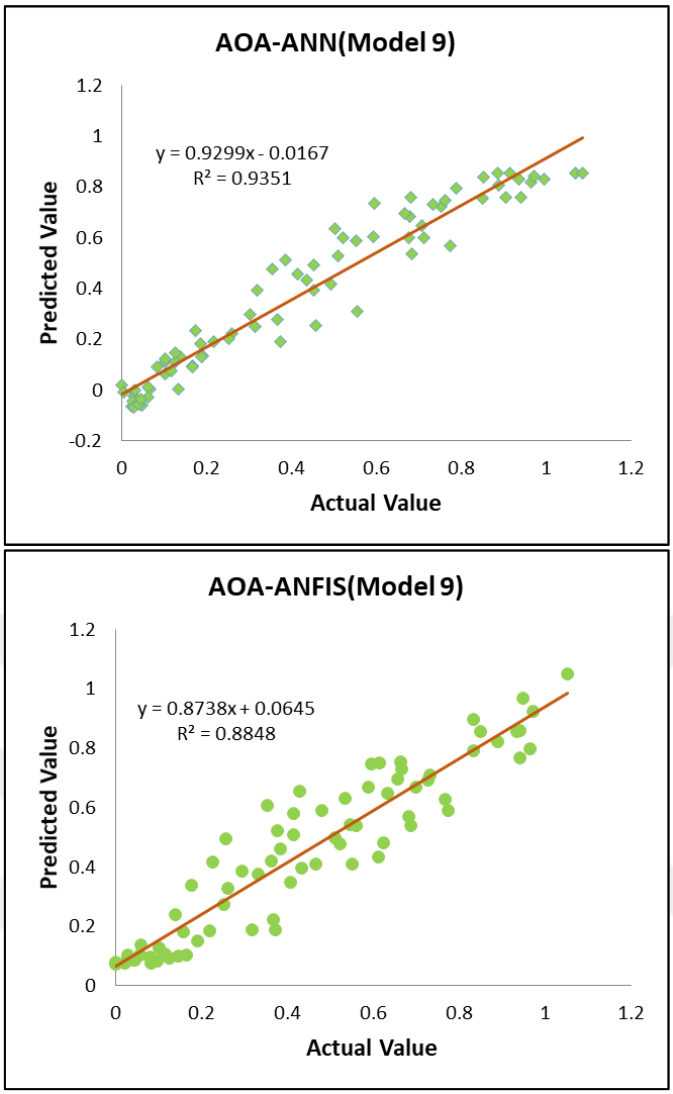


Figure 4.20 The graphical visualization (scatter plot) for the hybrid intelligence predictive model 9 (AOA-ANN &AOA-ANFIS).

In the realm of predictive modeling for Estimate at Completion, Taylor diagrams serve as invaluable tools for model comparison. These diagrams offer a concise visual representation of how well different predictive models perform in capturing the variability and correlation within EAC data. With their unique ability to display multiple metrics such as correlation coefficient, standard deviation ratio, and RMSE, Taylor diagrams enable analysts to assess the strengths and weaknesses of each model comprehensively. By leveraging the insights gleaned from Taylor diagrams, stakeholders can make informed decisions regarding which predictive model best aligns with the specific requirements and nuances of their EAC estimation processes, ultimately enhancing forecasting accuracy and project management efficiency.

In **Figure 4.21**, the Taylor diagram illustrates comparisons between the FFA-ANN and FFA-ANFIS models. As depicted in **Figure 4.21(a)**, the FFA-ANN prediction model incorporating Model 7 with seven input combination achieved the most accurate prediction closely aligning with the observed EAC values. Conversely, Model 6 emerged as the optimal choice for input variables in the FFA-ANFIS model, as highlighted in **Figure 4.21(b)**. Furthermore **Figure 4.22**, the Taylor diagram demonstrates comparisons made between the AOA-ANN and AOA-ANFIS models. Additionally, **Figure 4.22(a)** shows that - among the different AOA-ANN prediction models, Model 7 attained the most accurate prediction — which closely aligns with the observed EAC values. On the other hand, **Figure 4.22(b)** shows that Model 6 turned out to be the optimal choice for input variables in the AOA-ANFIS model.

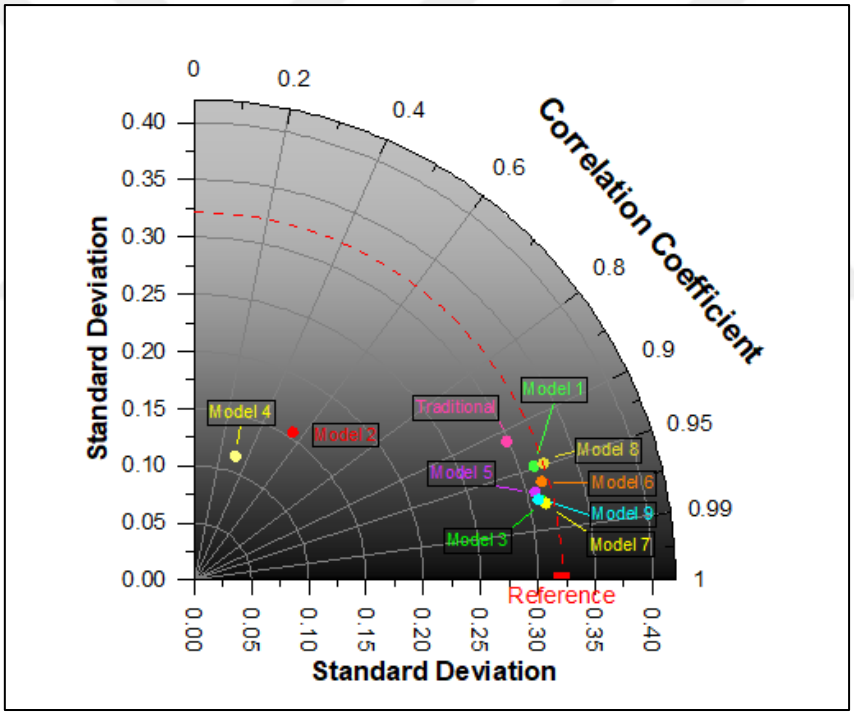


Figure 4.21(a): Taylor diagram for FFA-ANN Prediction Models

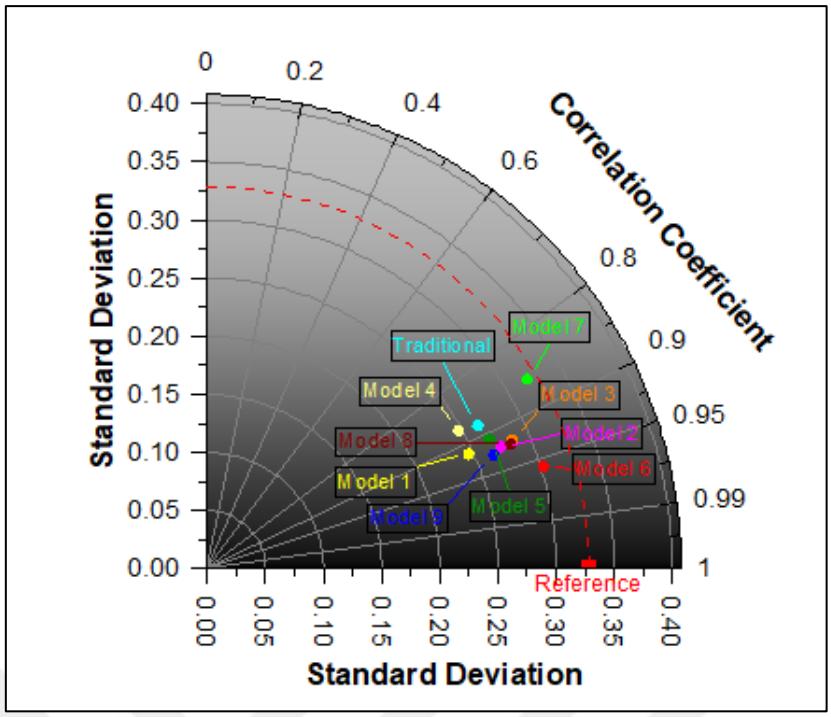


Figure 4.22(b): Taylor diagram for FFA-ANFIS Prediction Models

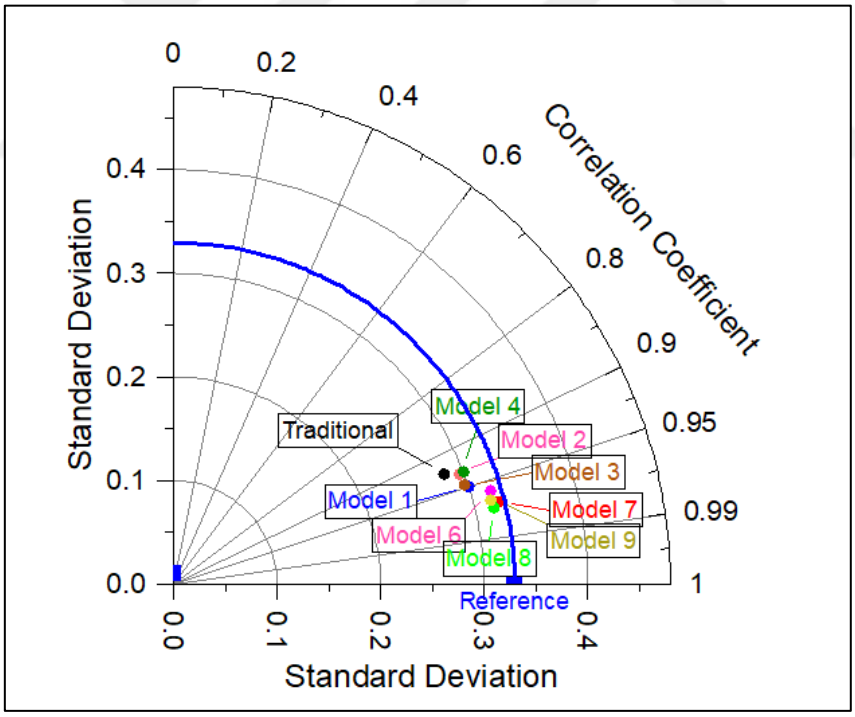


Figure 4.23 (a): Taylor diagram for AOA-ANN Prediction Models

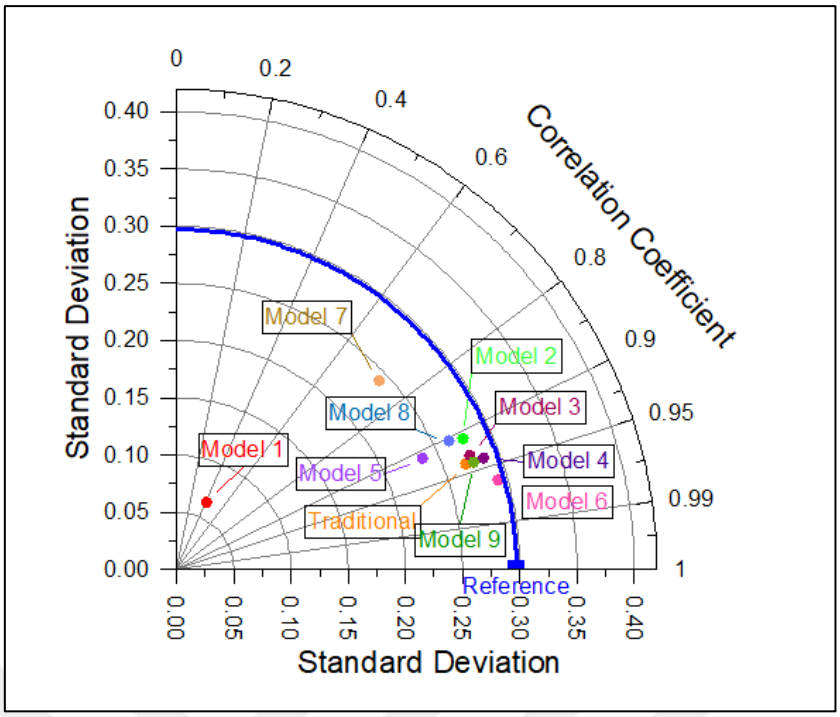


Figure 4.24 (b): Taylor diagram for AOA-ANFIS Prediction Models

The spider diagram (radar chart) is another way for model comparison, and it proves to be effective for comparing EAC prediction models. Moreover, the spider diagram allows analysts and stakeholders to simultaneously evaluate the strengths and vulnerabilities of a given model and across multiple criteria. **Figure 4.23** shows the comparison of the ANN, ANFIS and best models in AOA-ANN&AOA-ANFIS by using the MAE, RMSE, SMAPE, MSE and MRE the figure showing that AOA-ANN outperformed other models. Furthermore, the comparison of the best FFA-ANN and FFA-ANFIS models was shown in **Figure 4.24**. The figure showing that FFA-ANN outperformed FFA-ANFIS.

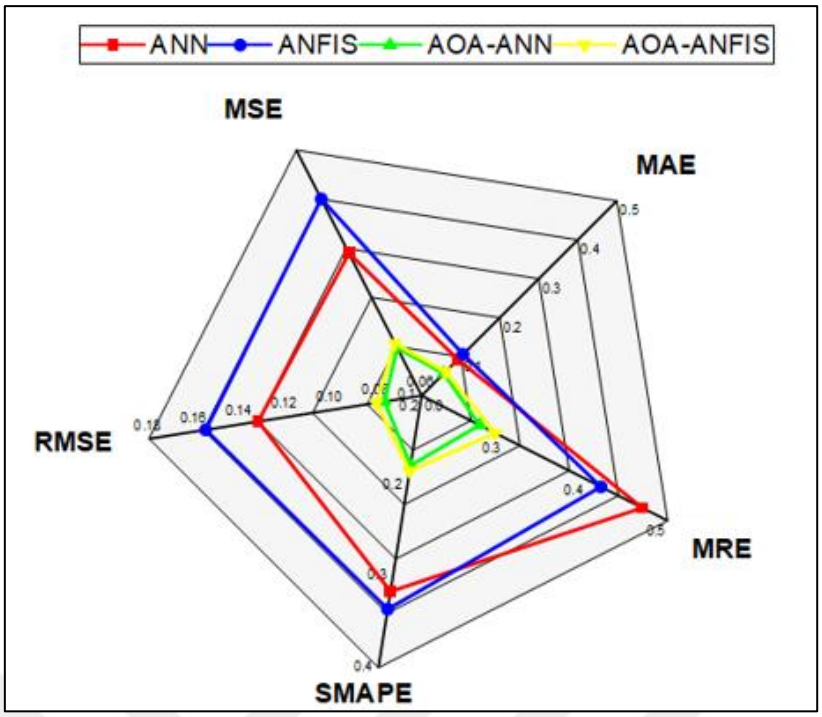


Figure 4.25: The Spyder diagram for ANN, ANFIS, and best models in AOA-ANN & AOA-ANFIS.

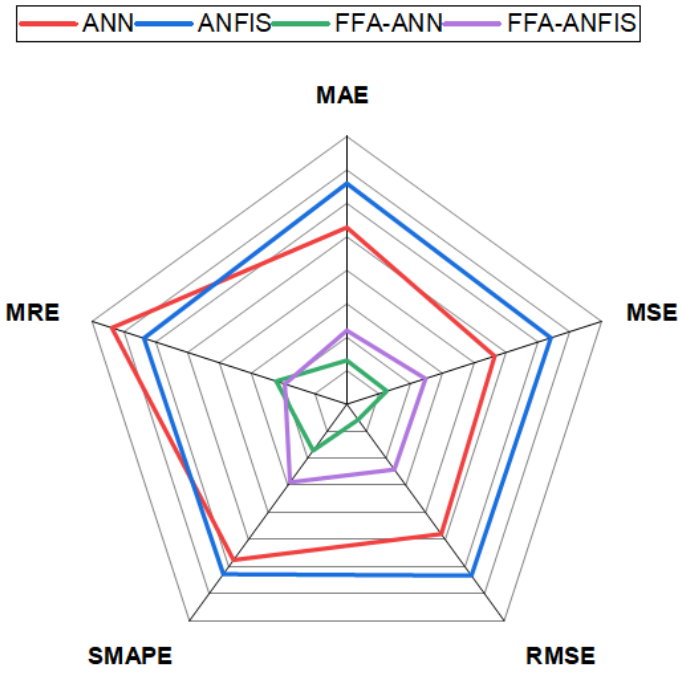


Figure 4.26: The Spyder diagram for ANN, ANFIS, and best models in AOA-ANN & AOA-ANFIS

5
CONCLUSION

Handling construction projects can be quite a puzzle, with uncertainties that can throw a wrench into the works. To tackle these uncertainties, budgets typically allocate resources like contingency and management reserves. These resources come into play to ensure that a project is not just completed but completed successfully. Now, what constitutes project success is a bit of a head-scratcher in project management discussions – there's a lot of talk, but not much consensus. Generally, though, it's about nailing the cost, time, and scope objectives, maintaining top-notch project management processes, and keeping the stakeholders happy. And that's exactly what this research is all about. We're diving into the world of new hybrid data-intelligence predictive models – FFA-ANFIS, AOA-ANFIS, FFA-ANN, and AOA-ANN – aiming to equip construction managers with a rock-solid methodology to control project costs and make spot-on estimations for the Estimate at Completion (EAC). This methodology is more than just theory; it's a fully automated system that lets you keep a keen eye on project activities, exercise control, and dodge any potential mishaps. The system has two key phases: first, the evolutionary phase, where FFA and AOA algorithms figure out which input attributes matter for the prediction matrix, and second, the ANN and ANFIS prediction models leveraging those chosen variables to model the EAC. To put the model through its paces in real-world scenarios, we've crunched the numbers from thirteen projects in the Taiwan region, showcasing the practical prowess of our intelligent system.

Initially, the findings solidified the predictability prowess of ANN compared to standalone ANFIS models. Furthermore, incorporating an input selection algorithm into the mix elevated the predictive performance. In quantitative terms, the optimal prediction accuracy was pinpointed, showcasing minimal absolute error metrics and an impressive fit-of-goodness (RMSE and correlation value (R) standing at (0.0686and 0.9767) with just 7 input parameters (CEI, Progress (%), COI, ACp,

SBI, CPI, CCI) harnessed by FFA-ANN. This underscored the capability of the proposed hybrid model to achieve dependable prediction accuracy even with a streamlined set of input variables.

The project faces limitations primarily rooted in the challenges of gathering data from construction projects, compounded by security and political issues. The policies of construction companies, intent on keeping project details confidential, further hinder data accessibility since the mistakes and errors that can introduce delays to projects often concealed from government or local authorities. Consequently, this research relies on historical data available in [12,13,14]. EAC in artificial intelligent models hinges on the input factors the model is constructed upon. Any alterations to these inputs can escalate estimated errors or render the model impractical, prompting the need for a dynamic linkage with cost index techniques to account for fluctuations in prices.

The prospects for future research hold significant relevance for the current study, paving the way for the implementation of this methodology in real-time applications across various construction projects. The potential contributions manifest as practical solutions, and it is imperative to underscore that the present investigation was conducted with a focus on residential construction projects, utilizing historical data. Exploring diverse construction projects, such as industrial endeavors, would serve as an exemplary extension, providing insights into more robust and project-specific monitoring capabilities. Addressing the uncertainties inherent in data, models, and input parameters remains a crucial aspect that warrants thorough examination in subsequent studies.

REFERENCES

[1] A. P. Kern and C. T. Formoso, “A model for integrating cost management and production planning and control in construction,” Journal of Financial Management of Property and Construction, vol. 11, no. 2, pp. 75–90, Aug. 2006, doi: 10.1108/13664380680001081/FULL/XML.

[2] J. Zeng, M. An, and N. J. Smith, “Application of a fuzzy based decision making methodology to construction project risk assessment,” International journal of project management, vol. 25, no. 6, pp. 589–600, 2007.

[3] S. A. Javed and S. Liu, “Bidirectional absolute GRA/GIA model for uncertain systems: application in project management,” IEEE Access, vol. 7, pp. 60885–60896, 2019.

[4] A. Mahmoudi and S. A. Javed, “Project scheduling by incorporating potential quality loss cost in time-cost tradeoff problems: The revised KKH model,” Journal of modelling in management, vol. 15, no. 3, pp. 1187–1204, 2020.

[5] M. Kasravi, A. Mahmoudi, and M. R. Feylizadeh, “A novel algorithm for solving resource-constrained project scheduling problems: a case study,” Journal of Advances in Management Research, vol. 16, no. 2, pp. 194–215, 2019.

[6] A. Mahmoudi and M. R. Feylizadeh, “A grey mathematical model for crashing of projects by considering time, cost, quality, risk and law of diminishing returns,” Grey Systems: Theory and Application, vol. 8, no. 3, pp. 272–294, 2018.

[7] M.-Y. Cheng, H.-C. Tsai, and E. Sudjono, “Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry,” Expert Syst Appl, vol. 37, no. 6, pp. 4224–4231, 2010.

[8] M.-Y. Cheng, N.-D. Hoang, A. F. V Roy, and Y.-W. Wu, “A novel time-depended evolutionary fuzzy SVM inference model for estimating construction project at completion,” Eng Appl Artif Intell, vol. 25, no. 4, pp. 744–752, 2012.

[9] D. S. Christensen, “Using performance indices to evaluate the estimate at completion,” The Journal of Cost Analysis, vol. 11, no. 1, pp. 17–23, 1994.

[10] E. F. T. AlHares and C. Budayan, “Estimation at completion simulation using the potential of soft computing models: Case study of construction engineering projects,” Symmetry (Basel), vol. 11, no. 2, p. 190, 2019.

[11] F. Nega, “Causes and effects of cost overrun on public building construction projects in Ethiopia,” Unpublished doctoral dissertation, Addis Ababa University, Ethiopia, vol. 34, 2008.

[12] F. M. Ottaviani and A. De Marco, “Multiple linear regression model for improved project cost forecasting,” Procedia Comput Sci, vol. 196, pp. 808–815, 2022.

[13] Y. Frimpong, J. Oluwoye, and L. Crawford, “Causes of delay and cost overruns in construction of groundwater projects in a developing countries; Ghana as a case study,” International Journal of project management, vol. 21, no. 5, pp. 321–326, 2003.

[14] M. R. Dube and S. K. Dixit, “Comprehensive measurement framework for enterprise architectures,” International Journal of Computer Science & Information Technology (IJCSIT), vol. 3, no. 4, pp. 71–92, 2011.

[15] F. T. Anbari, “Earned value project management method and extensions,” Project management journal, vol. 34, no. 4, pp. 12–23, 2003.

[16] Karrar Raoof Kareem KAMOONA, “THE IMPLEMENTATION OF HYBRID SOFT COMPUTING MODEL IN ESTIMATING PROJECT COMPLETION,” 2020.

[17] M.-Y. Cheng, H.-S. Peng, Y.-W. Wu, and T.-L. Chen, “Estimate at completion for construction projects using evolutionary support vector machine inference model,” Autom Constr, vol. 19, no. 5, pp. 619–629, 2010.

[18] A. F. V Roy, M.-Y. Cheng, and Y.-W. Wu, “Time dependent evolutionary fuzzy support vector machine inference model for predicting diaphragm wall deflection,” Neural Network World, vol. 24, no. 2, p. 193, 2014.

[19] C.-C. Huang and M.-Y. Cheng, “Estimate at completion for construction projects using evolutionary gaussian process inference model,” in 2011 International Conference on Multimedia Technology, IEEE, 2011, pp. 4414–4417.

[20] F. Giannuzzi, “Programme Management Monitoring and Controlling: statistical predictive models to improve Estimate at Completion,” 2023.

[21] D. Dvir, “Transferring projects to their final users: The effect of planning and preparations for commissioning on project success,” International Journal of Project Management, vol. 23, no. 4, pp. 257–265, 2005.

[22] M.-Y. Cheng, H.-C. Tsai, and E. Sudjono, “Evolutionary fuzzy hybrid neural network for conceptual cost estimates in construction projects,” in 26th International Symposium on Automation and Robotics in construction (ISARC2009), 2009, pp. 512–519.

[23] T. Narbaev and A. De Marco, “Combination of growth model and earned schedule to forecast project cost at completion,” J Constr Eng Manag, vol. 140, no. 1, p. 04013038, 2014.

[24] S. Vandevoorde and M. Vanhoucke, “A comparison of different project duration forecasting methods using earned value metrics,” International journal of project management, vol. 24, no. 4, pp. 289–302, 2006.

[25] D. S. Christensen, “Determining an accurate estimate at completion,” National Contract Management Journal, vol. 25, no. 1, pp. 17–25, 1993.

[26] D. S. Christensen, R. C. Antolini, and J. W. McKinney, “A review of estimate at completion research,” Cost Estimating and Analysis: Balancing Technology and Declining Budgets, pp. 207–224, 1992.

[27] P. A. Koushki, K. Al-Rashid, and N. Kartam, “Delays and cost increases in the construction of private residential projects in Kuwait,” Construction management and economics, vol. 23, no. 3, pp. 285–294, 2005.

[28] T. Narbaev and A. De Marco, “An earned schedule-based regression model to improve cost estimate at completion,” International journal of project management, vol. 32, no. 6, pp. 1007–1018, 2014.

[29] M.-Y. Cheng, H.-C. Tsai, and C.-L. Liu, “Artificial intelligence approaches to achieve strategic control over project cash flows,” Autom Constr, vol. 18, no. 4, pp. 386–393, 2009.

[30] Enas Fathi Taher ALHARES, “COUPLED INTELLIGENT PPREDICTIVE MODEL BASED GLOBAL HARMONY SEARCH AND EXTREME LEARNING MACHINE FOR MODELING PROJECT CONSTRUCTION ESTIMATION AT COMPLETION,” 2020.

[31] W. F. Abba, “Earned value management-reconciling government and commercial practices,” Program Manager, vol. 26, pp. 58–63, 1997.

[32] J. Pajares and A. Lopez-Paredes, “An extension of the EVM analysis for project monitoring: The Cost Control Index and the Schedule Control Index,” International Journal of Project Management, vol. 29, no. 5, pp. 615–621, 2011.

[33] A. Akintoye, “Analysis of factors influencing project cost estimating practice,” Construction Management & Economics, vol. 18, no. 1, pp. 77–89, 2000.

[34] G. R. Weckman, H. W. Paschold, J. D. Dowler, H. S. Whiting, and W. A. Young, “Using neural networks with limited data to estimate manufacturing cost,” 2010.

[35] M.-Y. Cheng, N.-D. Hoang, A. F. V Roy, and Y.-W. Wu, “A novel time-depended evolutionary fuzzy SVM inference model for estimating construction project at completion,” Eng Appl Artif Intell, vol. 25, no. 4, pp. 744–752, 2012.

[36] D. H. Curling, “Earned value project management: Quentin W Fleming and Joel M Koppleman Project Management Institute 1996 141 pp $24.95 (US) ISBN 1-880410-38-9,” 1998, Pergamon.

[37] T. Narbaev and A. De Marco, “An earned schedule-based regression model to improve cost estimate at completion,” International Journal of Project Management, vol. 32, no. 6, pp. 1007–1018, 2014.

[38] D. S. Christensen, “Using performance indices to evaluate the estimate at completion,” The Journal of Cost Analysis, vol. 11, no. 1, pp. 17–23, 1994.

[39] J. Hinze, Construction Planning and Scheduling, vol. 228. 2004.

[40] E. Kim, W. G. Wells Jr, and M. R. Duffey, “A model for effective implementation of Earned Value Management methodology,” International Journal of Project Management, vol. 21, no. 5, pp. 375–382, 2003.

[41] B. Kim and K. F. Reinschmidt, “Probabilistic forecasting of project duration using Bayesian inference and the beta distribution,” J Constr Eng Manag, vol. 135, no. 3, pp. 178–186, 2009.

[42] R. I. Carr, “Cost-estimating principles,” J Constr Eng Manag, vol. 115, no. 4, pp. 545–551, 1989.

[43] P. Sihag, F. Esmaeilbeiki, B. Singh, I. Ebtehaj, and H. Bonakdari, “Modeling unsaturated hydraulic conductivity by hybrid soft computing techniques,” Soft comput, vol. 23, pp. 12897–12910, 2019.

[44] C. Cha, GAO cost estimating and assessment guide: Best practices for developing and managing capital program costs. Government Printing Office, 2009.

[45] L. R. Dysert and T. W. Pickett, “So you think you’re an estimator?,” Cost engineering, vol. 47, no. 9, p. 30, 2005.

[46] R. E. Turochy, L. A. Hoel, and R. S. Doty, “Highway project cost estimating methods used in the planning stage of project development.,” Virginia Transportation Research Council (VTRC), 2001.

[47] C. Hendrickson and T. Au, Project management for construction: Fundamental concepts for owners, engineers, architects, and builders. Chris Hendrickson, 1989.

[48] K. B. Uppal, “Estimating? Past, present and 21st century,” AACE International Transactions, p. 118, 1997.

[49] I. ElSawy, H. Hosny, and M. A. Razek, “A neural network model for construction projects site overhead cost estimating in Egypt,” arXiv preprint arXiv:1106.1570, 2011.

[50] L. Dysert, “Is estimate accuracy an oxymoron?,” Cost Engineering, vol. 49, no. 1, p. 32, 2007.

[51] G.-H. Kim, S.-H. An, and K.-I. Kang, “Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning,” Build Environ, vol. 39, no. 10, pp. 1235–1242, 2004.

[52] T. M. S. Elhag, A. H. Boussabaine, and T. M. A. Ballal, “Critical determinants of construction tendering costs: Quantity surveyors’ standpoint,” International journal of project management, vol. 23, no. 7, pp. 538–545, 2005.

[53] P. F. Ostwald, “Construction cost analysis and estimating,” (No Title), 2001.

[54] Y.-R. Wang, C.-Y. Yu, and H.-H. Chan, “Predicting construction cost and schedule success using artificial neural networks ensemble and support vector machines classification models,” International Journal of Project Management, vol. 30, no. 4, pp. 470–478, 2012.

[55] M. Marjuki, “Computerised Building Cost Estimating System,” Doctoral dissertation, Universiti Teknologi Malaysia, 2006.

[56] A. Enshassi, S. Mohamed, and I. Madi, “Cost estimation practice in the Gaza Strip: A case study,” IUG Journal of Natural Studies, vol. 15, no. 2, 2015.

[57] C. M. Popescu, K. Phaobunjong, and N. Ovararin, Estimating building costs. Crc Press, 2003.

[58] K. K. Humphreys, Project and cost engineers’ handbook. CRC Press, 2004.

[59] R. E. Westney, The Engineer’s cost handbook: tools for managing project costs. CRC Press, 1997.

[60] AACE Recommended Practice and Standard, “Standard Cost Engineering Terminology,” 1990.

[61] Y. Li, H. Zhu, D. Wang, K. Wang, W. Kong, and X. Wu, “Comprehensive optimization of distributed generation considering network reconstruction based on Archimedes optimization algorithm,” in IOP conference series: earth and environmental science, IOP Publishing, 2021, p. 012031.

[62] M. Vanhoucke, Measuring time: Improving project performance using earned value management, vol. 136. Springer Science & Business Media, 2009.

[63] I. Henriques and P. Sadorsky, “The determinants of an environmentally responsive firm: An empirical approach,” J Environ Econ Manage, vol. 30, no. 3, pp. 381–395, 1996.

[64] J. L. Lollar, “Cost Performance Analysis Program for Use on Hand-Held Programmable Calculators,” Cost Research Report 141. Aeronautical Systems Division, Wright-Patterson …, 1980.

[65] C. W. Parker, “C/SCSC and C/SSR Cost Performance Analysis Programs,” DCAS Residency Harris, DCASMS, Orlando, Florida, vol. 20, 1980.

[66] J. B. Holeman, “A Product Improved Method for Developing A Program Management Office Estimated Cost at Completion,” Defense Systems Management School, Fort Belvoir, Virginia, 1975.

[67] J. A. Totaro, “A Logical Approach to Estimate at Completion Formulas,” Program Manager, vol. 16, pp. 29–33, 1987.

[68] J. Batselier and M. Vanhoucke, “Improving project forecast accuracy by integrating earned value management with exponential smoothing and reference class forecasting,” International journal of project management, vol. 35, no. 1, pp. 28–43, 2017.

[69] H. Khamooshi and A. Abdi, “Project duration forecasting using earned duration management with exponential smoothing techniques,” Journal of management in engineering, vol. 33, no. 1, p. 04016032, 2017.

[70] H. Khamooshi and H. Golafshani, “EDM: Earned Duration Management, a new approach to schedule performance management and measurement,” International Journal of Project Management, vol. 32, no. 6, pp. 1019–1041, 2014.

[71] W. Lipke, O. Zwikael, K. Henderson, and F. Anbari, “Prediction of project outcome: The application of statistical methods to earned value management and earned schedule performance indexes,” International journal of project management, vol. 27, no. 4, pp. 400–407, 2009.

[72] T. Scifers, “Software User’s Manual for Performance Analyzer Version 3.2,” 1991, Camp Springs, Maryland: TRW, Inc.

[73] J. Covach, J. J. Haydon, and R. O. Reither, “A study to determine indicators and methods to compute estimate at completion (EAC),” Virginia: ManTech international corporation, vol. 30, p. 39, 1981.

[74] H. R. Bright and T. Howard, “Weapon system cost control: Forecasting contract completion costs,” Redstone Arsenal, Alabama: Comptroller/Cost Analysis Division, United States Army Missile Command, 1981.

[75] M. Riedel and J. Chance, “Estimates at completion (EAC): A guide to their calculation and application for aircraft, avionics, and engine programs,” Aeronautical Systems Division, Wright-Patterson Air Force Base, OH, 1989.

[76] R. M. Skitmore and S. T. Ng, “Forecast models for actual construction time and cost,” Build Environ, vol. 38, no. 8, pp. 1075–1083, 2003.

[77] O. Swei, J. Gregory, and R. Kirchain, “Construction cost estimation: A parametric approach for better estimates of expected cost and variation,” Transportation Research Part B: Methodological, vol. 101, pp. 295–305, 2017.

[78] H. Li, Q. P. Shen, and P. E. D. Love, “Cost modelling of office buildings in Hong Kong: an exploratory study,” Facilities, vol. 23, no. 9/10, pp. 438–452, 2005.

[79] H. Watkins, “An Application of Rayleigh Curve Theory to Contract Cost Estimation and Control,” Naval Postgraduate School, Monterey, California, 1982.

[80] F. M. Ottaviani and A. De Marco, “Multiple linear regression model for improved project cost forecasting,” Procedia Comput Sci, vol. 196, pp. 808–815, 2022.

[81] G. M. Ayalew, M. G. Meharie, and G. G. Ayalew, “Regression modeling for prediction of earned value indexes in public building construction projects: the case of Ethiopia,” Cogent Eng, vol. 10, no. 1, p. 2220497, 2023.

[82] G. K. Chacko, “Improving Cost and Schedule Controls Through Adaptive Forecasting,” Concepts: The journal of Defense Systems Acquisition Management, vol. 4, pp. 73–96, 1981.

[83] M. Z. Fi-Sabban, “Forecast of Schedule/Cost Status Utilizing Cost Performance Reports of the Cost/Schedule Control Systems Criteria: A Bayesian Approach,” KJÜ, p. 105, 1974.

[84] M.-Y. Cheng, H.-C. Tsai, and W.-S. Hsieh, “Web-based conceptual cost estimates for construction projects using Evolutionary Fuzzy Neural Inference Model,” Autom Constr, vol. 18, no. 2, pp. 164–172, 2009.

[85] M.-Y. Cheng and A. F. V Roy, “Evolutionary fuzzy decision model for construction management using support vector machine,” Expert Syst Appl, vol. 37, no. 8, pp. 6061–6069, 2010.

[86] M. R. Feylizadeh, A. Hendalianpour, and M. Bagherpour, “A fuzzy neural network to estimate at completion costs of construction projects,” International Journal of Industrial Engineering Computations, vol. 3, no. 3, pp. 477–484, 2012.

[87] F. Caron, F. Ruggeri, and A. Merli, “A Bayesian approach to improve estimate at completion in earned value management,” Project Management Journal, vol. 44, no. 1, pp. 3–16, 2013.

[88] A. Enshassi, S. Mohamed, and S. Abushaban, “Factors affecting the performance of construction projects in the Gaza strip,” Journal of Civil engineering and Management, vol. 15, no. 3, pp. 269–280, 2009.

[89] I. P. A. Wiguna, S. Scott, and F. Khosrowshahi, “Nature of the critical risk factors affecting project performance in Indonesian building contracts,” in 21st annual ARCOM Conference, 2005, pp. 225–235.

[90] I. Mahamid and A. Bruland, “Cost deviation in road construction projects: The case of Palestine,” Australasian Journal of Construction Economics and Building, The, vol. 12, no. 1, pp. 58–71, 2012.

[91] V. T. Luu and S.-Y. Kim, “Neural network model for construction cost prediction of apartment projects in Vietnam,” 한국건설관리학회논문집, vol. 10, no. 3, pp. 139–147, 2009.

[92] D. Xiao-chen, “Application of neural network in the cost estimation of highway engineering,” J. Comput, vol. 5, no. 11, p. 1763, 2010.

[93] C. L. C. Roxas and J. M. C. Ongpeng, “An artificial neural network approach to structural cost estimation of building projects in the Philippines,” in Proceedings of the DLSU Research Congress, 2014, p. 1.

[94] K. R. Kareem Kamoona and C. Budayan, “Implementation of genetic algorithm integrated with the deep neural network for estimating at completion simulation,” Advances in Civil Engineering, vol. 2019, no. 1, p. 7081073, 2019.

[95] S. Kar, S. Das, and P. K. Ghosh, “Applications of neuro fuzzy systems: A brief review and future outline,” Appl Soft Comput, vol. 15, pp. 243–259, 2014.

[96] M. A. Awadallah, E. H. E. Bayoumi, and H. M. Soliman, “Adaptive deadbeat controllers for brushless DC drives using PSO and ANFIS techniques,” Journal of Electrical Engineering, vol. 60, no. 1, pp. 3–11, 2009.

[97] J. Zhao and B. K. Bose, “Evaluation of membership functions for fuzzy logic controlled induction motor drive,” in IEEE 2002 28th Annual Conference of the Industrial Electronics Society. IECON 02, IEEE, 2002, pp. 229–234.

[98] N. Talpur, M. N. M. Salleh, and K. Hussain, “An investigation of membership functions on performance of ANFIS for solving classification problems,” in IOP conference series: materials science and engineering, IOP Publishing, 2017, p. 012103.

[99] N. Walia, H. Singh, and A. Sharma, “ANFIS: Adaptive neuro-fuzzy inference system-a survey,” Int J Comput Appl, vol. 123, no. 13, 2015.

[100] M. Pratama, S. Rajab, and E. M. Joo, “Extended approach of ANFIS in cascade control,” International Journal of Computer and Electrical Engineering, vol. 3, no. 4, pp. 572–576, 2011.

[101] T. Uçar, A. Karahoca, and D. Karahoca, “Tuberculosis disease diagnosis by using adaptive neuro fuzzy inference system and rough sets,” Neural Comput Appl, vol. 23, pp. 471–483, 2013.

[102] X.-G. Wang, “Overview of the research status on artificial neural networks,” in 2nd International Forum on Management, Education and Information Technology Application (IFMEITA 2017), Atlantis Press, 2018, pp. 351–356.

[103] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks,” in Proceedings of the fourteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings, 2011, pp. 315–323.

[104] S. Liu, C. Zheng, and M. Liu, “Back propagation algorithm in feedforward netural network and its improvement,” Computer Science, vol. 23, no. 1, pp. 76–79, 1996.

[105] F. Bre, J. M. Gimenez, and V. D. Fachinotti, “Prediction of wind pressure coefficients on building surfaces using artificial neural networks,” Energy Build, vol. 158, pp. 1429–1441, 2018.

[106] D. Mitra and R. K. Paul, “Hybrid time-series models for forecasting agricultural commodity prices,” Model Assisted Statistics and Applications, vol. 12, no. 3, pp. 255–264, 2017.

[107] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, “Estimation of IMU and MARG orientation using a gradient descent algorithm,” in 2011 IEEE international conference on rehabilitation robotics, Ieee, 2011, pp. 1–7.

[108] M. Al-Zewairi, S. Almajali, and A. Awajan, “Experimental evaluation of a multi-layer feed-forward artificial neural network classifier for network intrusion detection system,” in 2017 International Conference on New Trends in Computing Sciences (ICTCS), IEEE, 2017, pp. 167–172.

[109] T. Ayer, Q. Chen, and E. S. Burnside, “Artificial neural networks in mammography interpretation and diagnostic decision making,” Comput Math Methods Med, vol. 2013, no. 1, p. 832509, 2013.

[110] K. Zor, O. Timur, and A. Teke, “A state-of-the-art review of artificial intelligence techniques for short-term electric load forecasting,” in 2017 6th international youth conference on energy (IYCE), IEEE, 2017, pp. 1–7.

[111] M. M. Mijwel, “Artificial neural networks advantages and disadvantages,” Mesopotamian Journal of Big Data, vol. 2021, pp. 29–31, 2021.

[112] M. G. M. Abdolrasol et al., “Artificial neural networks based optimization techniques: A review,” Electronics (Basel), vol. 10, no. 21, p. 2689, 2021.

[113] J. V. Abellan-Nebot and F. Romero Subirón, “A review of machining monitoring systems based on artificial intelligence process models,” The International Journal of Advanced Manufacturing Technology, vol. 47, pp. 237–257, 2010.

[114] D. Nauck and R. Kruse, “Obtaining interpretable fuzzy classification rules from medical data,” Artif Intell Med, vol. 16, no. 2, pp. 149–169, 1999.

[115] K. Salahshoor, M. Kordestani, and M. S. Khoshro, “Fault detection and diagnosis of an industrial steam turbine using fusion of SVM (support vector machine) and ANFIS (adaptive neuro-fuzzy inference system) classifiers,” Energy, vol. 35, no. 12, pp. 5472–5482, 2010.

[116] M. Abd Elaziz, A. A. Heidari, H. Fujita, and H. Moayedi, “A competitive chain-based Harris Hawks Optimizer for global optimization and multi-level image thresholding problems,” Appl Soft Comput, vol. 95, p. 106347, 2020.

[117] E. H. Houssein, M. R. Saad, F. A. Hashim, H. Shaban, and M. Hassaballah, “Lévy flight distribution: A new metaheuristic algorithm for solving engineering optimization problems,” Eng Appl Artif Intell, vol. 94, p. 103731, 2020.

[118] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," Future generation computer systems, vol. 97, pp. 849–872, 2019.

[119] F. A. Hashim, K. Hussain, E. H. Houssein, M. S. Mabrouk, and W. Al-Atabany, "Archimedes optimization algorithm: a new metaheuristic algorithm for solving optimization problems," Applied intelligence, vol. 51, pp. 1531–1551, 2021.

[120] A. Fathy, A. G. Alharbi, S. Alshammari, and H. M. Hasanien, "Archimedes optimization algorithm based maximum power point tracker for wind energy generation system," Ain Shams Engineering Journal, vol. 13, no. 2, p. 101548, 2022.

[121] X.-S. Yang, "Firefly algorithm, stochastic test functions and design optimisation," International journal of bio-inspired computation, vol. 2, no. 2, pp. 78–84, 2010.

[122] M. Heydari and S. Shabanlou, "A proposed novel hybrid intelligent model based on ANFIS integrated with firefly algorithm for forecasting discharge coefficient of side weirs on converging canals," Irrigation and Drainage, vol. 69, no. 4, pp. 865–879, 2020.

[123] P. N. Siva Jyothi and Y. Rohita, "A Review on PYTHON for Data Science," Machine Learning and IOT, International Journal of Scientific & Engineering Research, vol. 10, p. 12, 2019.

[124] V. Cutting and N. Stephen, "A review on using Python as a preferred programming language for beginners," Int. Res. J. Eng. Technol, vol. 8, no. 8, pp. 4258–4263, 2021.

[125] D. Lakshminarayanan and S. Prabhakaran, "A study on python programming language," Dogo Rangsang Res. J, vol. 10, pp. 2347–7180, 2020.

[126] I. Attarzadeh and O. S. Hock, "Implementation and Evaluation of Earned Value Index to Achieve an Accurate Project Time and Cost Estimation and Improve 'Earned Value Management System,'" in 2009 International Conference on Information Management and Engineering, IEEE, 2009, pp. 312–316.

[127] K. I. Payne, "An Investigation of the Stability of the Cost Performance Index," Air Force Institute of Technology, Wright-Patterson AFB, Ohio, 1990.

[128] W. Lipke, O. Zwikael, K. Henderson, and F. Anbari, "Prediction of project outcome: The application of statistical methods to earned value management and earned schedule performance indexes," International journal of project management, vol. 27, no. 4, pp. 400–407, 2009.

[129] S. Kim, "Project success indicators focusing on residential projects: Are schedule performance index and cost performance index accurate measures in earned value?," Canadian Journal of Civil Engineering, vol. 36, no. 11, pp. 1700–1710, 2009.

[130] K. M. Nassar, H. G. Gunnarsson, and M. Y. Hegab, "Using Weibull analysis for evaluation of cost and schedule performance," J Constr Eng Manag, vol. 131, no. 12, pp. 1257–1262, 2005.

[131] R. K. Cox, “Managing change orders and claims,” Journal of Management in Engineering, vol. 13, no. 1, pp. 24–29, 1997.

[132] C. Charoenngam, S. T. Coquinco, and B. H. W. Hadikusumo, “Web-based application for managing change orders in construction projects,” Construction Innovation, vol. 3, no. 4, pp. 197–215, 2003.

[133] V. S. A. Kargi and A. Öztürk, “Subcontractor selection using analytic hierarchy process,” Business and Economics Research Journal, vol. 3, no. 3, p. 121, 2012.

[134] M. J. T. Camilleri, Implications of climate change for the construction sector: Houses. BRANZ, 2000.

[135] N. El-Sawalhi and M. Mahdi, “Influence of climate change on the lifecycle of construction projects at Gaza strip,” Journal of Construction Engineering and Project Management, vol. 5, no. 2, pp. 1–10, 2015.

[136] T. P. Williams, “Predicting changes in construction cost indexes using neural networks,” J Constr Eng Manag, vol. 120, no. 2, pp. 306–320, 1994.

[137] Z. M. Yaseen et al., “Novel approach for streamflow forecasting using a hybrid ANFIS-FFA model,” J Hydrol (Amst), vol. 554, pp. 263–276, 2017.

[138] J. Zhou, C. Li, C. A. Arslan, M. Hasanipanah, and H. Bakhshandeh Amnieh, “Performance evaluation of hybrid FFA-ANFIS and GA-ANFIS models to predict particle size distribution of a muck-pile after blasting,” Eng Comput, vol. 37, no. 1, pp. 265–274, 2021.

[139] H. Riahi-Madvar, M. Dehghani, R. Memarzadeh, and B. Gharabaghi, “Short to long-term forecasting of river flows by heuristic optimization algorithms hybridized with ANFIS,” Water Resources Management, vol. 35, pp. 1149–1166, 2021.

[140] M. Ehteram et al., “Performance improvement for infiltration rate prediction using hybridized adaptive neuro-fuzzy inferences system (ANFIS) with optimization algorithms,” Ain Shams Engineering Journal, vol. 12, no. 2, pp. 1665–1676, 2021.

[141] L. Zhang, J. Wang, X. Niu, and Z. Liu, “Ensemble wind speed forecasting with multi-objective Archimedes optimization algorithm and sub-model selection,” Appl Energy, vol. 301, p. 117449, 2021.

[142] R. Mustafa, P. Samui, and S. Kumari, “Reliability analysis of gravity retaining wall using hybrid ANFIS,” Infrastructures (Basel), vol. 7, no. 9, p. 121, 2022.

PUBLICATIONS FROM THE THESIS

Conference Papers

1. Abo Mhady A., Gürgün A. P., Budayan C. (2023). Cost Estimate At Completion using Novel Hybrid Intelligent Model Based on Artificial Neural Network (ANN) Integrated with Firefly Algorithm (FFA). 12. Uluslararası Akademik Araştırmalar Kongresi (pp. 126–136).

Papers

1. Mhady, A. A., Budayan, C., & Gurgun, A. P. (2024). Estimate-at-completion (EAC) prediction using Archimedes optimization with adaptive fuzzy and neural networks. *Automation in Construction*, *166*, 105653.