REPUBLIC OF TURKEY

YILDIZ TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

THE IMPLEMENTATION OF HYBRID SOFT COMPUTING MODEL IN ESTIMATING PROJECT COMPLETION

Karrar Raoof Kareem KAMOONA

DOCTOR OF PHILOSOPHY THESIS

Department of Civil Engineering

Civil Engineering (English) Program

Advisor

Assoc. Prof. Dr. Cenk BUDAYAN

August, 2020

REPUBLIC OF TURKEY

YILDIZ TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

THE IMPLEMENTATION OF HYBRID SOFT COMPUTING MODEL IN ESTIMATING PROJECT COMPLETION

A thesis submitted by Karrar Raoof Kareem KAMOONA in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY is approved by the committee on 19/08/2020 in Department of Civil Engineering, Civil Engineering (English) Program.

Assoc. Prof. Dr. Cenk BUDAYAN
Yildiz Technical University
Advisor

Approved By the Examining Committee

Assoc. Prof.Dr. Cenk BUDAYAN, Advisor
Yildiz Technical University

Assoc. Prof. Dr. Zeynep IŞIK, Member
Yildiz Technical University

Assoc. Prof. Aslı Pelin GÜRGÜN, Member
Yildiz Technical University

Prof. Dr. Beliz ÖZORHON, Member
Bogazici University

Prof. Dr. Hüsnü Murat GÜNAYDIN, Member
Istanbul Technical University

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 The Implementation of Hybrid Soft Computing Mode in Estimating Project Completion supervised by my supervisor, Assoc. Prof. Dr. Cenk BUDAYAN. In the case of a discovery of false statement, I am to acknowledge any legal consequence.

Karrar Raoof Kareem KAMOONA

Signature

ACKNOWLEDGEMENTS

Consequent to expressing profound gratitude to God for giving the strength and capacity to finish this research, I might need to show the truthful to the family who have conferred in me the drive and support to complete this work.

The completion of this study would not have been possible without the assistance and incentives from some deferential individuals to whom I might want to appreciate:

I am grateful to thesis supervisor and advisor Assoc. Prof. Cenk BUDAYAN for his invaluable guidance, encouragement, availability and continuous support throughout the course of Doctorate program.

Special thanks to my friends for their support always, Mother, brothers and sisters.

The endeavors of the greater part of the teachers and the college instructors who liberally shared their insight in various phases of my past training are additionally appreciatively valued.

Karrar Raoof Kareem KAMOONA

TABLE OF CONTENTS

LIST OF SYMBOLS

X	Input
Y	Output
$\varphi(x_1)$	Presents the high order of the feature space
ξ	Slack Variable
f	The activation function
b	The bias
C	The positive regularization
w	The weight matrix
M	Training data set

LIST OF ABBREVIATIONS

ACWP	Actual Cost of Work Performed
ACWP	Actual Cost of Work Performed
AI	Artificial Intelligence
BAC	Budget at Completion
BCWP	Budgeted Cost of Work Performed
BCWS	Budgeted Cost of Work Performed
BF	Brute Force
CCI	Construction Price Fluctuation
CPI	Cost Performance Index
CPR	Contract Performance Report
CV	Cost Variance
DNN	Deep Neural Network
EAC	Estimates at Completion
ECD	Estimated Completion Date
EFHNN	Evolutionary Fuzzy Hybrid Neural Network
ES	Earned Schedule
EV	Earned Value
EVM	Earned Value Management
EVMS	Earned Value Management System
FHNN	Fuzzy Hybrid Neural Network
FL	Fuzzy Logic
GA	Genetic Algorithm
GP	Gaussian Process
IEAC	Independent Estimates at Completion
IPMR	Integrated Program Management Report
MAE	Mean Absolute Error
MMAE	Multiple Model Adaptive Estimation
MRE	Mean Relative Error

NAVWESA	Navy Weapons Engineering Support Activity
NN	Neural Network
NSE	Nash Sutcliffe Coefficient
PCD	Planned Completion Date
RE	Relative Error Percentage
RMSE	Root Mean Square Error
SI	Scatter Index
SPI	Schedule Performance Index
SRA	Schedule Risk Assessment
SV	Schedule Variance
SVM	Support Vector Machine
SVR	Support Vector Regression
TSARISM	Time Series Analysis for Army Internal Systems Management
WI	Wilmot's Index

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The Implementation of Hybrid Soft Computing Model in Estimating Project Completion

Karrar Raoof Kareem KAMOONA

Department of Civil Engineering

Doctor of Philosophy Thesis

Advisor: Assoc. Prof. Dr. Cenk BUDAYAN

In construction project management, there are several factors influencing the final project cost. Among various approaches, estimate at completion (EAC) is an essential approach utilized for final project estimation. The main merit of EAC is including the probability of the project performance and risk. In addition, EAC is extremely helpful for project managers to define and determine the critical problems throughout the project progress and determine the appropriate solutions to these problems. In this research, a relatively new intelligent model called deep neural network (DNN) is proposed to calculate the EAC. The proposed DNN model is authenticated against one of the predominated intelligent models conducted on the EAC prediction, namely support vector regression model (SVR). In order to demonstrate the capability of the model in the engineering applications, historical project information obtained from fifteen projects in Iraq region is inspected in this research. The second phase of this research is about the integration of two input optimization algorithms hybridized with the proposed and the comparable predictive intelligent models. These input optimization algorithms are genetic algorithm (GA) and brute force algorithm (BF). The aim of integrating these input optimization algorithms to approximate the input attributes and investigate the highly influenced factors on the calculation of EAC. Overall, the enthusiasm of this study is to provide a robust intelligent model that estimates the project cost accurately over the traditional methods. Also, the second aim is to introduce a reliable methodology that can provide efficient and effective project cost control. The proposed GA-DNN is demonstrated as a reliable and robust intelligence model for EAC calculation.

Keywords: Estimate at completion, cost project management, deep neural network, support vector regression

ÖZET

Tahmini Proje Tamamlama Maliyetinin Tahmin Edilmesinde Hibrit Yumuşak Hesaplama Modelinin Uygulanması

Karrar Raoof Kareem KAMOONA

İnşaat Mühendisliği Anabilim Dalı

Doktora Tezi

Danışman: Doç. Dr. Cenk BUDAYAN

İnşaat proje yönetiminde, nihai proje maliyetini etkileyen birçok faktör vardır. Çeşitli yaklaşımlar arasında, tahmini tamamlama maliyeti (TTM), nihai proje tahmini için kullanılan temel bir yaklaşımdır. TTM'nin esas değeri, proje performansının ve risk olasılığını içermesidir. Ek olarak, TTM proje yöneticileri için proje ilerleyişindeki kritik sorunları tanımlamak ve belirlemek ve bu sorunlara uygun çözümleri belirlemek için son derece yararlıdır. Bu araştırmada, TTM'yi hesaplamak için derin sinir ağı (DSA) olarak adlandırılan nispeten yeni bir akıllı model önerilmiştir. Önerilen DSA modeli, TTM tahmin edilmesinde kullanılan baskın akıllı modellerden birine, yani destek vektör regresyon modeline (VRM) dayanarak doğrulanmıştır. Modelin mühendislik uygulamalarındaki kabiliyetini göstermek için Irak bölgesinde on beş projeden elde edilen tarihi proje bilgileri bu araştırmada incelenmiştir. Bu araştırmanın ikinci aşaması önerilen ve karşılaştırılabilir akıllı modellerin iki giriş optimizasyon algoritmasının entegrasyonu ile hibritlendirilmesi ile ilgilidir. Bu girdi optimizasyon algoritmaları genetik algoritma (GA) ve kaba kuvvet algoritmasıdır (BF). Bu girdi optimizasyon algoritmalarının entegre edilmesindeki amaç ise girdi niteliklerine yaklaştırmak ve TTM hesaplamasında son derece etkili faktörleri araştırmaktır. Genel olarak, bu çalışmanın hedefi, geleneksel yöntemlere göre proje maliyetini doğru tahmin eden güçlü bir akıllı model geliştirmektir. Ayrıca, ikinci amaç ise verimli ve etkin proje maliyet kontrolü sağlayabilecek güvenilir bir metodoloji sunmaktır. Önerilen GA-DSA, TTM hesaplaması için güvenilir ve sağlam bir tahmin modeli olduğu bu çalışmada gösterilmiştir.

Anahtar Kelimeler: Tahmini tamamlama maliyeti, proje maliyet yönetimi, derin sinir ağı, destek vektör regresyonu, girdi yaklaşımı, hibrit akıllı model

---

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

1
INTRODUCTION

1.1 Background

In the construction industry, project success has always foundered on the high uncertainty in an operational environment. Thus, it is not surprising that construction projects frequently suffer cost overrun [1]. To operate profitably, construction companies must frequently evaluate project cost at completion to detect deviations and to carry out appropriate responses. However, construction firms typically focus on budget planning during the initial project stage, which practically ignores the impact of engineering cost changes and information updates during construction [2]. This fact prevents effective project cost control and detection of potential problems. Therefore, cost estimation is a crucial task, and it needs to be carried out at various stages of a project [3]. Moreover, the accuracy of construction cost estimation is a critical factor in the success of the project [4]. Poor cost estimation may result in profit loss and occasionally leads to project failure.

The building industry is deemed one of the most dangerous and diverse markets. It does, therefore, play a significant role in the creation and achievement of social goals. It is one of the main sectors and leads in developed countries to about 10 per cent of the gross national product (GNP) [5]. In building projects, Ajayi [6] noticed that the most important success metrics for ensuring that the project was carried out appropriately are: the consistency of the job, execution of the project on schedule, the contractor's profitability rate and completion of the project within the projected budget. Many of investigated pieces of literature on construction projects suggested that the common criteria for project management success are generally considered to cost, time and quality [7–11]. In a study conducted by Ajayi [12] to highlight the factors affecting the performance of contractors on construction projects within Lagos State, Nigeria form contractors' perceptions through a questionnaire survey. It was concluded that the most significant factors are: site conditions, the complexity of a project, lack of subcontractors experience, communication between project parties, lack of labours' experience and poor material quality. Long et al. [13] grouped the problems of construction projects performance into five groups: incompetent designers/ contractors, poor management skills, social and technological issues, site-related issues, and improper techniques and tools.

Poor planning and lack of communication between project parties were concluded by many researchers as key factors negatively affecting performance in construction projects [14–17]. Sweis et al. [18] conducted a study to address the main causes of poor contractor performance on a construction project in Jordan. They found that the top leading causes are: financial difficulties faced by the contractor, manpower shortage and excessive change orders.

Project management is the application of tools, techniques, processes, and knowledge skills to manage and control a unique, temporary, and multidisciplinary task [19]. Project management can be seen as a process that starts from the definition of project scope and objectives and ends with the fulfilment of project requirements. As the Project Management Institute (2008) stated, the main phases involved in project management are: initiating, planning, executing, monitoring and controlling, and closing. An important process that requires the project team to plan and control the complexity of the activities involved in the project is the calculation of estimation at completion (EAC), both in terms of cost and time. Project management requires a forward control mechanism to manage the high level of uncertainty and innovation that can affect a project, and in this context estimates at completion represent a prerequisite for identifying and implementing suitable corrective measures on time.

Selecting the appropriate methodology that accurately estimates the cost at completion of ongoing projects based on current performance is an open issue in the area of project monitoring. A largely used technique which has been showing itself as an objective and valuable tool to accurately compute cost EAC is the conventional Earned Value Management (EVM) methodology with associated Cost Performance Index (CPI) and Schedule Performance Index (SPI). However, the technique has some limitations in calculating the cost EAC such as failing to address a critical path, probabilistic estimations, work quality and risk impact. Thus, the application of statistical methods and simulation techniques has been considered as an attempt to overcome the limitations inherent with conventional index based EAC.

Earned value (EV) has evolved from a hot globe acquisition process topic to managerial best practice over the years [20]. Earned Value Management (EVM) is an assemblage of business management practices which provides a structured method of performance measurement and analysis, and as such, is not considered as a software program [21]. EV measures can help project managers to be aware of the status of their programs in terms of cost and performance categorisation if properly applied and analysed. The EVM system (EVMS) serves as a way of organisational project scheduling, budgeting, and component planning to predict and forecast future status. With the EVMS, program managers can predict the results during a decision-making process. Project is often altered to meet the established goals; such alterations stem from a comparison of the present state of a program to the predicted measure. The accuracy of the predicted estimations determines the efficiency of a decision-making process. The EVMS methodology provides the inputs (termed Estimates at Completion (EAC)) that highlights the measures for these forecasts.

Some of the undesirable program outcomes (like cost overruns, cancellations and schedule delays) are products of a poor decision-making process due to inaccurate estimates [10,22]. Calcutt in 1993 observed that nearly (20-50%) of all completed projects were over-budgeted in terms of phase and type [23]. Additionally, projects with 15 to 20% cost overruns are unlikely to be completed with a reduced cost overrun [24]. Some of the known program failures, like the Navy's A-12 Avenger program, contributed to the list of management problems that compelled the Department of Defense (DoD) to sponsor the search for more accurate project estimation methods [25].

A major problem today in the acquisition of new weapon systems is cost control due to the declining defence budgets and the rapid advancements in technology [26]. A study conducted by the RAND Corporation observed about 20% cost growth (excluding the effects of inflation and increased purchase quantity) in the 197 defence programs it investigated [27]. The major outcome of the study was "that there is a systematic downward bias in cost estimates which has been resulting in cost growth". Furthermore, the study implied that the techniques used for the calculation of the final costs of contracts for defence acquisition (called EACs) are not only accurate but also underestimate the actual final costs of such contracts.

Recent examples of control costs failure include the Navy's A-12 program (cancelled in Jan. 1991) and the Air Force's C-l 7 program (faced persistent congress criticism) [28]. The reason for the cancellation of the Navy's A-12 program was due to the lack of a good estimate of the expected final cost of the program.

A major contributor to this problem is the existence of several ways of EAC calculation. Such methods range from the simple index-based methods to complicated statistical methods. Several reviews have been conducted on this prospect of computing projects estimations at completion. McKinney presented 47 formulas for the calculation of EAC based on 18 sources [29]. Additionally, several complicated nonlinear regression-based techniques, such as the Rayleigh probability distribution coupled with Multiple Model Adaptive Estimation (MMAE), Rayleigh probability distribution, and a modified Beta distribution [30,31] [32], [33] have been proposed by other researchers. The existence of numerous methods made it difficult to select the best one to be used.

Regrettably, most of the comparative studies have concentrated on the index-based and regression methods. Despite the advent of the new advanced soft computing models, there is still a challenge of solving this problem, especially when several parameters are manipulating the EAC value.

The research is primarily aimed at using a deep neural network model to determine the complex nonlinearity, non-stationarity and stochasticity relationship between the independent variables (i.e., schedule performance index (SPI), subcontractor billed index, cost variance (CV), schedule variance (SV), cost performance index (CPI), owner billed index, construction price fluctuation (CCI), change order index, climate effect index) and the corresponding EAC (the dependent variable).

1.2 Research Scope and Contribution

In project management, on-time delivery within budget is a fundamental factor that highlights the importance of monitoring how a project is doing. Performance measurement is the ongoing, regular collection of information that can provide this controlling system. Evaluation is a specific, in-depth way to gather and analyse information and come to conclusions about the performance of the policy, program or strategy at the project level. Therefore, forecasting the completion time of the project during its execution can play a significantly important role to evaluate and control a project. There are two general types of evaluations:

“Formative” or “process” evaluations that are designed to improve the design and implementation of the program, policy or strategy as it unfolds, and

“Summative” or “outcome” evaluations that are designed to judge a program, policy or strategy’s relevance, success and/or cost-effectiveness (including its relative contribution to the intended outcomes).

Nowadays, a plethora of business activities is managed as projects. This has stimulated, in recent years, a growing interest in project management techniques, developed to foresee and manage project evolution with greater care. The study of project management techniques started in the early 1950s intending to provide managers with a powerful tool to identify the critical path for projects. This effort was extended later to include project risk analysis and management. It became apparent that considering the tasks as given and predetermined could lead to misleading results in project management practice. In real projects, it is usually inadequate to define a good schedule based on deterministic processing times, because these times are only estimates and are susceptible to unpredictable changes.

Project risks originate from the uncertainty that is present to a different extent in all projects. A significant amount of work has been done to conceptualise and measure uncertainty. Projects are subject to considerable uncertainty due to several possible sources. Resources may become unavailable, activities duration may experience some delay, new activities may be incorporated in the project, or other activities may be even deleted. Amongst the full range of sources of significant uncertainty associated with any given project [34], an obvious aspect of uncertainty concerns estimates of potential variability of activity duration. Our choice in this context involves modelling processing times of activities as random variables. We should mention that, while sometimes the type of uncertainty encountered in particular classes of projects could not fit the axiomatic basis of probability theory (examples can be drawn from projects with highly non-repetitive nature). In most cases, probability-based methods are well suited to this kind of problem and have been successfully applied for decades in the project management literature.

In this study, the EAC was simulated based on several related input variabilities (i.e., schedule performance index (SPI), subcontractor billed index, cost variance (CV), schedule variance (SV), cost performance index (CPI), owner billed index, construction price fluctuation (CCI), change order index, climate effect index). The applied methodology for the relationship determination was based on the artificial intelligence (AI) concept. Diverse versions of AI models have introduced for solving the related problem of the EAC such as adaptive neuro-fuzzy inference system (ANFIS), artificial neural network (ANN), support vector regression (SVR), genetic programming (GP) and several others. However, due to the high complexity of the problem that is associated with nonlinearity, non-stationarity and stochasticity relationship between the independent variables, there were several limitations indicated over the literature based on the developed AI models. Among several AI drawbacks, the requirement for the internal parameters' optimisation, 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 new version of AI models called deep neural network (DNN) model to determine the EAC. The data used for the analysis was restricted to construction contracts executed in Iraq due to time constraints. 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.

In response to some limitation of the explored AI models and the significant advancement in the theoretical and technological capabilities, deep learning has emerged and is rapidly expanding as one of the most exciting fields of science. It is being used in technologies such as self-driving cars, image recognition on social media platforms, and translation of text from one language to others. Deep learning is the subfield of machine learning that is devoted to building algorithms that explain and learn a high and low level of abstractions of data that traditional machine learning algorithms often cannot. The models in deep learning are often inspired by many sources of knowledge, such as game theory and neuroscience, and many of the models often mimic the basic structure of a human nervous system. As the field advances, many researchers envision a world where software isn't nearly as hardcoded as it often needs to be today, allowing for a more robust, generalised solution to solving problems [35].

1.3 Estimation at Completion (EAC) Concept

Organisations and customers often require project managers to authenticate the possibility of meeting the cost and schedule of their projects within the approved budget, the Budget at Completion (BAC), and the Planned Completion Date (PCD). The EAC and the Estimated Completion Date (ECD) serve as measures towards answering these questions. As per the EV guidelines, EAC is defined as the cumulative sum of the contracts' Actual Cost of Work Performed (ACWP) and the project manager's best estimate of the time-phased resources required to actualise the remaining authorised work (ETC). Often, this relationship is expressed thus:

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

Hence, the EAC can be referred to as the prediction of the final project cost. The project manager may revisit work priorities; the remaining project tasks may be replanned or reschedule to ensure the completion of the project's goals within the estimated budget. The major concern is to complete the project within the budget and schedule.

In all estimates, there are always changes in the level of uncertainty of an EAC based on the nature of the remaining task, information at hand, and the perceived remaining risks. To ensure prudent project management, there is a need to know the EACs' validity, especially when there is a significant variation of the EAC from the approved project budget. Therefore, project management mainly aims at the identification of the degree of uncertainty associated with the remaining project tasks, determining the cost of the remaining project task, and managing the influence of uncertainty on the project cost goals.

As such, three separate EACs are required in the Contract Performance Report (CPR) and the Integrated Program Management Report (IPMR) in a bid to capture information on the level of cost uncertainty or the degree unexpected project risks. EACs that represents the worst-case (or highest potential cost), the best case (or lowest potential cost), and the most likely EAC (the project manager's best estimate) are required in these reports to forecast the project completion time (ECD).

Being that the actual cost-to-date is a known value, the uncertainty of EAC is dependent on the EAC itself. The preparation of the ETC requires the re-estimation of the required resources to complete the remaining part of the approved work using the cost experience to date before applying a range of factors like the current direct and overhead rates, Monte Carlo simulations, Schedule Risk Assessment (SRA), root cause analysis, etc.

A properly planned ETC also considers anticipated labour rate, labour efficiency, purchase order commitments, price and usage of materials, other direct cost prices and usage performance, resources by type, risk and opportunities, and other factors identified by the top management. Furthermore, the ETC should also be mapped to the current schedule and must be consistent with the ECD.

A mathematical or independent estimate of EAC is often prepared (as a way of cross-checking the EAC) based on the cost and schedule experience using current performance indices. For instance, the Cost Performance Index (CPI) and Cumulative Budgeted Cost for Work Performed (ACWP) can be used to complete the EAC by dividing the project BAC by the CPI. The type of EAC resulting from such computation is often called Independent EAC (IEAC) and is different from the formal or grassroots EAC as it can be easily prepared and used to validate the feasibility of the current cost estimate and to indicate when to undertake a comprehensive EAC. It should be noted that these computations do not consider any thought of the anticipated labour rate and efficiency, risk and opportunities, SRA, etc. They are often said to be independent of sanity, judgment, and logic but are computed from comparative analysis.

It is becoming an industrial practice to regularly prepare EACs along with the computation of the Best Case, Worst Case, and Most Likely EACs. At least, a comprehensive EAC (a complete "bottoms-up" EAC) is annually required on projects that are subject to the requirements of the Defence Federal Acquisition Regulation DFAR 252.234-7002 Earned Value Management System. A comprehensive EAC is also often prepared before initialising any major project as it can reflect the reduced uncertainty resulting from the release of project design and bill of material. The comprehensive EAC helps the project manager to provide the answer to the following questions:

- Can the remaining authorised funds be enough to complete the project?
- Can the future cost performance be predicted from the previous cost experience?
- Is it possible to modify the remaining project based on the performance to date?
- Is there any impact of the project cost performance on the corporate financial condition?

Hence, a timely release of the EAC and ECD must be an integral aspect of the project and financial management practices since both demands a continuous comparison of the EAC and ECD with the project targets to predict the actual financial performance for both the stockholders and customers.

1.4 Research Objectives

Based on the inspiration of developing a reliable and robust predictive model, it is imperative to design a fast and effective system which considers the issues of cost control during the project execution for the prediction of project EAC by using AI methods. The aim of this study rallies on resolving the identified issues in project cost management through the collection of relevant historical data and studies about project cost management for the identification of the factors that significantly affect project cost. As a developing country, Iraq population grows and increasing demand for housing. The historical data are collected from several residential construction projects located in the Iraq region from some private companies in addition to the local authority in Baghdad. This project information is used to set up the trend of a project cost flow and the relationship between project EAC and monthly costs were mapped based on historical knowledge and experience. Based on historical data, a new intelligent model called deep neural network (DNN) model is developed for the prediction and control of EAC variation during project execution. The suggested DNN model validated against support vector regression (SVR) prediction model. The second phase of the current research devoted to the implementation of a hybrid evolutionary model called a genetic algorithm (and brute force) integrated with deep neural network GA-DNN (and BF-DNN). The aim of applying the evolutionary phase as a prior stage for the predictive model is to allocate the correlated attributes to build an accurate predictive model. Again, the modelling of the hybrid intelligent model is authorised with the GA-SVR and BF-SVR. This step ensured that the identification of potential issues for effective measures to be timely implemented.

1.5 Thesis Outlines

The objectives of this study are met in the following manner. The first chapter of this PhD thesis provides a comprehensive introduction of the research which identifies the research motivations, objectives and outline of the thesis. Apart from the introduction in chapter 1, the thesis is composed of the following chapters.

Chapter 2 presents an extensive literature review of previous studies conducted on estimation at the completion of various engineering projects. This chapter started with EAC traditional calculations. Then after, the chapter presented an extensive state-of-the-art focus on the researches have been conducted over the past two decades and covers the latest advances pure artificial intelligence/complementary models as well as an assessment of the surveyed studies.

In chapter 3, the proposed deep neural network is presented. Besides, the benchmark data-driven model is described (i.e., support vector regression). The input optimisation algorithms are also briefly explained. Finally, the case study dataset and modelling development are described.

Chapter 4 showed the results of the modelling applications. There are several performance indicators applied to examine the predictive models. A comprehensive discussion and analysis for the application carried out including the efficiency of the models, the performance of individual and hybrid models, and how much wellness the proposed method in capturing the nonlinearity of the studied problem.

Finally, chapter 5 presents the findings and contributions of this research and proposes several efforts for future research in the domain of projects management modelling and particularly on estimation at completion.

2
LITERATURE REVIEW

2.1 Introduction

In nature, project management is often anticipated with non-avoidable risks. The constant environmental changes and other external constraints have made risk management a serious issue in the construction industry [36,37]. Project monitoring must be given adequate attention, (in terms of the close monitoring and detection of deviations and of taking appropriate measures to address any deviations) to make a profit. Meanwhile, the initial stage of most construction activities focusses on budget planning, effectively neglecting the impact of changes in the engineering cost and the updating of information during construction [38], and this has prevented an effective detection of the problems associated with project cost control. Owing to the dynamic nature of project conditions upon the commencement of a project, there is a need for a regular revision of the project budget for effective project execution.

The importance of early planning to final project outcomes is emphasized widely in the literature [37,39]. However, these plans generally cannot be applied entirely, and they are revised throughout the project. Therefore, a constantly reviewed plan is required for effective project management which must reflect the actual condition of the project. Thus, the required actions can be taken when the project is going out of the control. Otherwise, cost overruns are noticed towards the end of the contract, and at this stage, remedial approaches may be ineffective and late. However, during the early phase of projects, the construction companies usually focus on the budget planning and generally ignore the other areas such as changes in cost, information update and cost management [38].

Although cost control is very important in the construction projects, the cost control is a time consuming and difficult process, since several factors, which can affect the cost of projects, and the influences of these factors should be considered individually at each stage of the project [2]. Estimate at completion (EAC) is one of the important indicators to perform cost control [40–42]; however, it should be calculated accurately to identify the problems and develop appropriate responses. Earned Value Management (EVM) is one of the widely used methods among project managers to calculate EAC [20,21,43]. In this methodology, project cost, schedule and scope metrics are integrated into a single measurement system to measure and analyze actual project status against its baseline and estimate the project total cost and duration at completion [44]. A brief description of the EAC process during the life of a project is displayed in Figure 2.1.

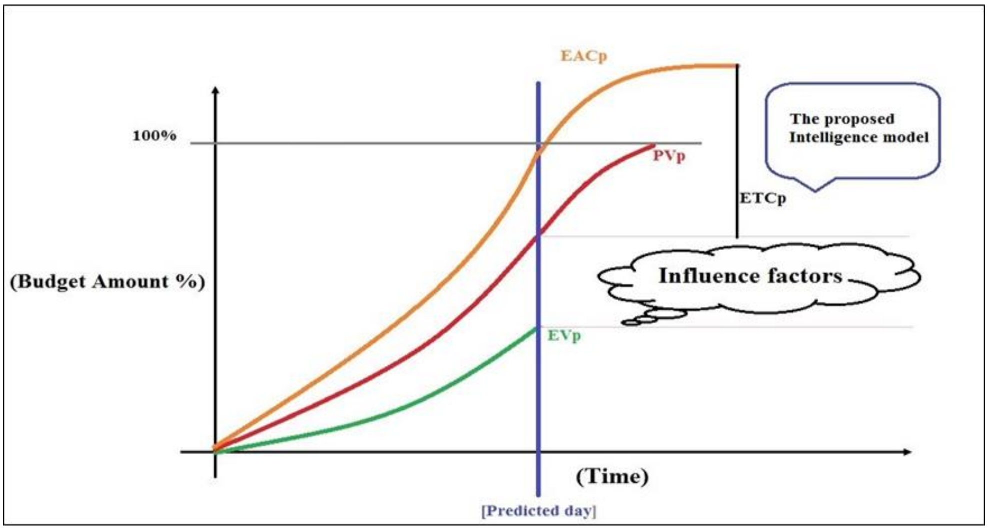


Figure 2.1 The proposed hybrid intelligent model for the prediction of the estimation at completion (EAC) [45].

It has been stated that the EVM is gaining broader acceptance owing to increase the acknowledgement of its capability of reducing the problems of EVM also improving utilities [46]. However, traditional EVM methods have some limitations. For example, the EVM methods developed by using index-based methodology are suffering from the usage of just past information and performance index in the calculation of the remaining budget [47]. Besides, these models provide unreliable cost forecasts in the early stages of project life due to the limited number of EVM data [20]. Although accurate predictions were achieved with the traditional EVM methods when used on some special projects, there were obvious errors in most of the cases. This has led to an industrial situation of not knowing the right prediction approach to be selected for a project. Another drawback of the EVM is that revisions must be manually conducted as it is applied to each process of a construction project, thereby making EVM a complicated and time-consuming method.

Consequently, it is critical to computerize engineering management process if EVM is to be deployed effectively for project cost control. Construction companies mostly engage computer systems which are reasonably powerful in the analysis of the initial stage of construction budgets. The computer system cannot respond to changes at each stage of construction or predict construction project EAC using the EVM method. Hence, the development of using artificial intelligence (AI) models to solve this problem is extensively attractive for the new era of research trend for project engineering scholars.

2.2 Indexed Formulation for EAC

The cost management data which the contractor provided in the Cost Performance Report or the Cost/Schedule Status Report can be used to formulate the EAC [48]. İn this chapter, the reviewed studies presumed the reliability of data presented in these reports. Data reliability is dependent on the degree a contractor adheres to an internal control system which involves the budgeting, scheduling, and analysis of contractual effort.

All formulas for EAC estimation depend on the integration of numerous data parameters contained in the cost management report. These data parameters include the Budgeted Cost of Work Scheduled (BCWS); Actual Cost of Work Performed (ACWP), and Budgeted Cost of Work Performed (BCWP). The BCWS, ACWP, and BCWP are often reported every month. The cumulative/average data is then computed throughout the contract duration [28].

The formulas for EAC estimation are grouped into artificial intelligence modelling, index, and regression methodologies. Equation (2.1) represents the generic index-based formula.

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

Where "c" represents the cumulative data. The BAC represents the total budget approved for the identified project.

The index method, which is often a combination of ACWP, BCWP, and BCWS, is mainly used for adjusting the cost of the remaining project work (BAC - $BCWPc$). The adjustment is mainly based on the assumption that the past cost and schedule performance of the project reflects the future performance. There are four groups of performance indices:

Cost Performance Index: $CPI = BCWP/ACWP$ (2.2)

Schedule Performance Index: $SPI = BCWP/BCWS$ (2.3)

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

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

The weights (W_1 and W_2) shown in equation (2.5) can assume any value in the range of 0 to 1 and normally add to unity.

These indices can be based on either monthly, average, or cumulative data and for the labelling, the following conventions are used: "CPI_m" = CPI based on the most recent month; "CPI_c" = cumulative CPI; "CPI_x" = CPI averaged over x number of months, starting from the most recent month to the later. For instance, CPI_3 represents a three months average CPI calculated from the current month and the immediate past two months. The same conventions are used for SPI and SCI; for instance, "SPI_6" represents a six months average SPI calculated from the current and the immediate past five months. There are two ways to average these indices; the averaged index is usually expressed as a ratio of sums through x months [48]:

$$CPI_x = \sum BCWP_x/ACWP_x \tag{2.6}$$

$$SPI_x = \sum BCWP_x/BCWS_x \tag{2.7}$$

Alternatively, it can be defined by dividing the sum of the monthly indices by the number of months [48]:

$$CPI_x = \frac{\sum CPI_m}{x} \tag{2.8}$$

$$SPI_x = \frac{\sum SPI_m}{x} \tag{2.9}$$

In this paper, an averaged index is defined according to Equations (2.6) and (2.7) unless otherwise specified.

The 2nd and 3rd groups of EAC formulas are called "regression" and "other." Either linear or nonlinear regression analysis is used to derive the regression-based EAC formulas. Nonlinear regression analysis is defined in this paper as nonlinear relationship analysis irrespective of the possibility of transforming it into a linear relationship 3. In most cases, the ACWP is often the dependent variable while BCWP, a performance index, or time is the independent variable. The "other" category comprised of formulas that fall outside of the first two categories, such as the heuristics-based formulas.

There are several formulas for the estimation of EAC, and this has placed the analyst in the position of deciding the right formula to be used. The Performance Analyzer is a well-known analysis software which allows the user to select from a range of formulas. However, there is no guideline on how to select the best or accurate formula. This issue will be addressed in the remaining parts of this work by reviewing the previous studies conducted on EAC in the last sixteen years.

2.3 Empirical Formulation for EAC

The empirical formulations for EAC describe a "new" modelling methodology which generally does not involve statistical or sophisticated heuristic technique but cannot be fully described as comparative analysis.

Index-based methods: A proposal for four empirical methods based on the composite index weights has been made (Jakowski, 1977; Lollar, 1980; Parker, 1980; Totaro, 1987). The formulation of these methods depends on the subjectively assigned index weights. These studies believe that the information content of SPI is eventually lost since it is driven to unity at contract completion. As the project nears completion, the weight assigned to the SPI tends to decrease to zero. The fifth study by Haydon derived a point estimate from a range of EACs estimated by several index-based formulas [49].

A complicated heuristic method for the determination of the composite index weights was proposed by Jakowski (Navy Aviation Systems Command, 1977). The method first used CPIc until significant decreases are noted in the most recent monthly-based CPIs. Once this is noted, an "optimally weighted" composite index will be used. An optimal weighting is taken to be the exact weight that resulted in the least historical standard deviation in the composite index. As the project approaches 60 % completion, CPIc is used again. Although the original report for Jakowski's heuristic is not available, it has been described by [50].

The definition of the weights for cumulative SPI and CPI was proposed by Lollar (Aeronautical Systems Division, 1980) as the contribution of the values of the cost and schedule variance percentages to their total [51]. Blythe later adopted this definition (1982) and Cryer (1986) in their comparative studies. However, it performed poorly against the other formulas [52,53].

A method proposed by Parker (Defense Logistics Agency, 1980) involves the computation of a range of composite indices by varying the weights in the range of 0 to 1 with increments of 0.1. Based on the conditions of the project, the analyst will decide the most appropriate composite index to be used.

With a suitable variable transformation, the general linear regression model can apply to inherently linear models. For instance, it is possible to transform nonlinear cumulative cost growth patterns (which sometimes, are closely approximated by logistics curves) into linear forms before their estimation using ordinary least squares.

Totaro (Defense Logistics Agency 1987) suggested that the determination of the composite index weights should depend on the percentage completion [54]. The analyst assigns starting weights for the SPI and CPI after considering the characteristics of the program, such as the projected manpower loading by the contractor.

A technique for the development of a point estimate from a range of EACs estimated with different formulas [49] has been proposed by Haydon and Riether (ManTech Corporation for Navy Weapons Engineering Support Activity (NAVWESA), 1982). First, index-based formulas are used to compute a range of EACs [50] before expanding the range by 2.5 %. The median of the expanded range is considered as the point estimate for the EAC. After the analysis of 21 completed or almost completed projects handled by the Navy (15 development and six production projects), contractors with EAC value less than this point estimate are said to have a more accurate forecast most times. Then, a sample procedural worksheet and a numerical example are provided.

2.4 Regression-based Model for EAC Computing

Regression-based methods: A proposal for three non-comparative studies of modelling the curvilinear cumulative cost growth profile typical on defence contracts using regression analysis has been made. The proposed methods in these studies are complicated and well documented but require the knowledge of regression analysis and as such, are not easy to be implemented.

The forecasting of EAC using time series analysis has been proposed by Sincavage (Army Aviation Systems Command, 1974). The "Time Series Analysis for Army Internal Systems Management" (TSARISM) is a computer-based model which uses either moving average, autoregressive or a combination of both methods. This model is sensitive to autocorrelation (a statistical problem) and requires data collected over several months to be developed. As such, the model is only useful in the later stages of a project. According to the author of this model, the original documentation is no longer available.

The time-series forecasting method used by the B-1 System Program Office has been described by Olsen et al. (Aeronautical Systems Division, 1976) [55]. İn this system, the EAC was predicted using the “GETSA”, a computer program developed by General Electric and leased by the B-1 SPO. A brief description of other techniques such as exponential smoothing and regression analysis was also provided. A numerical example was also provided.

An alternative approach to the development of nonlinear regression-based model coefficients has been described by Busse (Air Command and Staff College, 1977) [56]. Although this description did not make any comparisons with the Karsch model [57], it provided a numerical example using Karsch data. A comparison of the results of Busse and Karsch at certain stages of project completion showed the EACs generated by the Karsch model to be more accurate.

The fitting of data from development programs into a normalized S-curve using nonlinear regression analysis has been proposed by Weida (Air Force Academy, 1977) [58]. After the adjustment of the data for statistical and inflation problems, Weida stated that the S-curve described the cumulative cost growth on all the examined 22 development programs. The normalized S-curve can then be employed for both predictive and comparative tasks. The study provided a numerical example. Despite the complicated nature of Weida’s technique, it is compelling and deserves serious consideration.

The use of a time-series forecasting method called "Adaptive Forecasting" has been proposed by Chacko (Defense Systems Management College, 1981). This method requires five months of data before making an accurate EAC estimate [59]. The adaptive forecasting model essentially changes as the data for each month becomes available. This model is best used for short-term EAC forecasting.

The prediction of EAC using linear regression analysis and an adaptive form of the Rayleigh-Norden model has been proposed by Watkins (Navy Postgraduate School, 1982) [30]. Watkins proposed the use of the Rayleigh-Norden model for the description of the life-cycle patterns of manpower buildup and phaseout on defence contracts. The model is used in this study to perform a linear regression analysis of ACWP against time using quarterly data from 3 contracts submitted to C/SSRs after adjusting the data for inflation. There was no adjustment for autocorrelation.

The calculation of EAC using a Bayesian probability theory has been proposed by El-Sabban (Army Aviation Systems Command, 1973) [60]. İn this method, a normal probability distribution, a mean, and a variance are assumed for EAC at the beginning of a project. With the availability of the current data on ACWP, the Bayes's formula is used to reverse the "previous probability distribution" of the EAC. This model is more useful in the initial phase of a project since it is not dependent on a long history of performance data. Although the accuracy of the method has been challenged, it is still well presented [61] with the provision of an example.

The development of EAC using a "performance factor" which is determined by a subjective decision as a "product improved method" has been proposed by Holeman (Defense Systems Management College, 1974) [62]. The performance factor is used as a performance index and includes a linear combination of several variables. İt is the duty of the analyst to determine the relative contribution of each variable. Holeman also suggested the subjective determination and evaluation of a range of EACs using simulation. The method presented a numerical example.

The conventional EVM index-based methods for cost estimation at the completion of an ongoing project are characterized with limitations which are inherent with both early-stage unreliability and the belief that past EVM data is the best available information. In a bid to address such limitations, a new CEAC methodology, which based on a modified index-based formula. A modified index-based formula has been proposed for the prediction of the expected cost for the remaining part of a project. This method employs the Gompertz growth model through a nonlinear regression curve fitting [43]. The proposed method relies on schedule progress as a determinant of cost performance, as its equation is integrated with an Earned Schedule-based factor which indicates the expected duration for project completion. The model is suited for the early, middle, and late-stage estimates compared to the four previously compared index-based formulae. The method is developed as a practical tool which helps project managers to make better incorporation of the progress status into the task of CEAC computation. İt also extended EVM studies by providing a better understanding of the relationship between cost and schedule factors.

A new regression-based nonlinear CEAC method which is a combination of a growth model and earned schedule (ES) concepts have been proposed to improve the accuracy of the early forecasting of the final cost at the completion of an ongoing project [44]. This method provides the CEAC for projects at the early and mid-completion stages. Hence, the research is formulated based on three major objectives; (i) to develop a new formula which is based on the combination of the ES method with four different growth models (Logistic, Gompertz, Bass, and Weibull). (ii) to validate the new method by applying it on nine past projects; and (iii) to select the best-performing growth model based on the test of their statistical validity and comparing their CEAC estimation accuracies. The statistical validity analysis of the four growth models and the comparison of their CEAC errors showed the Gompertz model-based CEAC formula to provide better fitting. And generate more accurate final cost estimates compared to the estimated computed by the index-based method and the three other models. Theoretically, the proposed method contributed to the EAC study by the combination of EV metrics with regression-based studies. It also has a practical implication of using an accurate and viable forecasting method which considered schedule impact as a factor that determines the cost behaviour.

2.5 Soft Computing Models for EAC

Several elements play essential roles in the construction of project feasibility studies; such elements include the basis of project evaluation, conceptual cost estimates, engineering design, cost management, and cost budgeting. İn the practical perspective, construction cost estimates are generally reliant on the intuitive experience of the expert. For raising the accuracy of conceptual cost estimates, there is a need to develop and employ scientific methods during project design and planning. The improvement of the accuracy of cost estimates based on a new evolutionary fuzzy neural network has been proposed [63]. The study combined the advantages of genetic algorithms, neural networks, and fuzzy logic into one intelligent predictive model. This intelligent predictive model is highly useful in the identification of optimal solutions for complex problems. The authors concluded the study by providing two types of estimators which can be used to make a precise estimate of conceptual construction cost in the early phase of projects.

Due to the challenging nature of construction projects, construction decision-makers have been compelled to successfully work in an environment characterized by frequent complexity and full of uncertainties. Being that numerous decisions are intuitively made based on limited data, two factors are mainly involved in making a successful decision; these are the experience of the concerned expert(s) and the quality of accumulated knowledge from past experiences. However, the value and accuracy of knowledge can be influenced by several factors. The evolutionary fuzzy support vector machine inference model was proposed as an alternative method towards utilizing and retaining knowledge gained from experience [64]. The study defined two actual and simulated construction management problems and showed the proposed hybrid intelligent model as an effective method of solving several construction-related problems.

During project feasibility studies, conceptual cost estimates and its impact on the final project success are important. This is because such estimates provide important information which can help in the evaluation of the project, drawing of the engineering designs, planning, and management of the budget cost. A hybrid evolutionary fuzzy hybrid neural network (EFHNN) model was proposed by [65] for the improvement of the precision of conceptual cost estimates. İn this model, a hybrid neural network (HNN), is first formed by combining neural networks (NN) with high order neural networks (HONN). This hybrid network operates with alternating linear and nonlinear neuron layer connectors. Then, fuzzy logic (FL) is integrated into the HNN to handle uncertainties; this gives rise to a fuzzy hybrid neural network (FHNN). Since a genetic algorithm is used in the FL and HNN to optimize the hybrid FHNN, the final version deployed in this work may be rightly called an 'EFHNN'. İn this study, we calculated and compared the estimates of the overall and category costs for actual projects. From the results, the proposed EFHNN can be used as an accurate cost estimator at the initial phase of projects. However, the linear and nonlinear neuron layer connectors in the EFHNN performed better than models that use a singular linear NN.

Project managers use EAC to project the total cost of a project at its completion because it is an important tool used to monitor the performance and risk of a project. The executives often make high-level project decisions, but they may lack the technical knowledge, and this may cause decision errors. The estimation of EAC using an evolutionary predictive model based on Gaussian process (GP) integrated with particle swarm optimization algorithm (PSO) has been proposed by [45]. The model proposed in this study was combined with an AI method and used to simulate human decision-making pattern in handling management problems. In this model, the input-output relationship was determined by the GP, while PSO was used as an optimization tool to optimize the hyper-parameters in the data function. The possibility of using the proposed hybrid model in EAC determination and the calculation of the chances of a change in the predicted model monitor was investigated in this study. The model served as a reliable EAC estimator and can be of help to project managers when there is a need to improve project cost control effectiveness. The results achieved from the study were validated on real data from construction projects.

EVM is commonly used as an effective control method for the detection of project status and forecasting of estimation at completion (EAC) cost. Normally, the conventional EAC prediction methods extend the current project situation to the future by using a prior performance factor. The prediction of EAC and future event cost using a fuzzy neural network model has been performed by Feylizadeh et al. [66]. The fuzzy approach combined both qualitative and quantitative factors that affect EAC prediction. The effectiveness of this model was investigated in a real-world case study, and the preliminary outcome of the investigation was encouraging.

During the construction phase of projects, cost overruns are frequently encountered; hence, there is a need for a proactive way of project costs monitoring and detecting potential problems. EAC is used in construction management as an indicator for helping project managers identify potential challenges and develop necessary solutions. In this study, the distinct characteristics in EAC prediction were handled using a weighted support vector machine model (WSVM), fuzzy logic, and fast messy genetic algorithm [40]. The features of time series data were addressed in this study by employing the WSVM model as a supervised learning technique. The aim of developing the fuzzy logic model was to enhance the capability of the model in approximate reasoning and to handle the uncertainties in the prediction of EAC. From the simulation studies, the developed model was found to achieve a significant enhancement in EAC modelling.

Using a Bayesian approach based on expert opinion can permit the exploitation of subjective judgments rigorously and formally; this can lead to an improved estimate at completion accuracy within the EVM framework. This model was exploited by Caron et al. [67] as a combination of experts' knowledge with project data records in a bid to create a future-oriented and more valuable support tool that allows forecasting process improvement. This has a direct influence on the associated improvement in the decision-making process concerning the project control. The proposed model in this study is robust and can be applied in every phase of a project, particularly during the early phase of a project when data records are unreliable. The model also allows the confidence interval estimation that describes future scenarios likely to be faced in the project. The effectiveness and applicability of the Bayesian approach were demonstrated and certified in the oil and gas industry. The results of the study also point out that the same approach can be applied to other industries to improve the processes of project control. For improving the input and output management strategy in the Bayesian model and make it friendlier, the model has been translated into a software package.

In Gaza strip, the project parties (owners, consultants, and contractors) were asked to fill a questionnaire that was developed to address the factors affecting contractor performance in construction projects. The survey was performed by Enshassi, Mohamed, and Abushaban [68] and they found that the main contributors to poor performance in construction project in Gaza: borders closure, lack of resources, leadership skills, change in material prices, lack of experience, poor equipment and poor materials quality. Mbachu and Nkando [69] found that quality and attitude to service is a main factor leading to project management success in South Africa. Wiguna and Scott [70] conducted a study determine the factors of cost and time overrun in construction projects in Indonesia through a questionnaire survey. In 22 building construction projects under concern in their study, the contractors agreed on the top affecting factors, namely: high inflation, design changes, defective design, weather conditions and payments delay. Conflict, poor workmanship and incompetence of contractors were found to be among the factors leading to client dissatisfaction in South Africa [71]. Eighty-four contractors from the West Bank in Palestine were asked by Mahamid [72] to identify the main factors affecting contractors' business failure through a questionnaire survey. Forty-four factors were considered in the study and were listed under three groups: (1) financial, (2) managerial and (3) external. He concluded that the top affecting factors are escalation in construction material cost, payments delay, lack of experience, low margin of profit and political situation.

Frimpong, Oluwoye, and Crawford [10] concluded the following factors to be among the main factors affecting time and cost performance in construction projects in

Ghana: financial difficulties faced by contractor, poor management and poor technical performances. Koushki, Al-Rashid, and Kartam [73] investigated the time and cost performance construction projects in Kuwait. Four hundred and fifty private residential project owners and developers have been randomly selected for personnel interview. They concluded that the main contributors to poor time and cost performance are related to owners, namely: financial difficulties and lack of experience. In order to improve the performance of construction projects, they suggested the followings: (1) availability of adequate funds for project owner, (2) more attention should be paid for time and cost planning at the early stages of the project, (3) improve tender selection methods. To enhance performance in construction projects, Omran, Abdalrahman, and Pakir [74] recommended improving the leadership skills of project team and leader.

A new study was developed in 2017, which focused on the use of ANN to establish an automated method for the estimation of the duration of dam projects construction [75]. In the study, the existing literature was reviewed while expert interviews were conducted on the critical variables that influence dam projects execution. Several data were collected on Iranian dam projects and used for the development of 7 ANN models. The performance was assessed using different datasets. The performance was assessed in terms of the root mean square error and correlation values which served as indicators for models' validity and reliability. To allow the stakeholders to use the ANN method to estimate the duration of dam project completion, a web-based automated prototype of the model was developed and validated.

By taking into account of the itemized limitations of the methodologies introduced for the EAC simulation, it is imperative to design a fast and effective system which considers the issues of cost control during the project execution for the prediction of project EAC using AI techniques. The aim of this study depends on resolving the identified issues in project cost management through the collection of relevant historical data and studies about project cost management for the identification of the factors that significantly affect project cost. The historical data are collected from several construction projects located in the Iraq region. The study information was used to set up the trend of a project cost flow and the relationship between project

EAC and monthly costs were mapped based on historical knowledge and experience. Based on historical data, a new artificial intelligence model called deep neural network (DNN) model is developed for the prediction and control of EAC variation during project execution. The suggested DNN model validated against one of the predominant prediction models (i.e., support vector regression) model. The second phase of the current research devoted to the implementation of a hybridized predictive model by integrating the DNN and SVR models with genetic algorithm (GA) and brute force algorithm (BF) optimization algorithms. The aim of applying those input selection phase as a prior stage for the predictive model is to allocate the correlated attributes to build an accurate predictive model.

3
METHODOLOGY

3.1 Introduction

Since the creation of the ANN in the 1940s, it has experienced much progress. Computational models of NN have been in existence for over five decades now, starting with the simple model developed by McCulloch and Pitts in 1943 [76]. The work of Donald Hebb on Hebbian learning allowed researchers to apply this knowledge to computational models in the late 1940s.

Frank Rosenblatt created a perception in 1958 as a type of binary classifier for pattern recognition which is based on a 2-layer computer learning network [77]. However, the work of Stewart and Watson, which identified issues in the existing computational neural models stagnated the field of ANNs until the development of multi-layered perception [78].

Paul Werbos developed the backpropagation algorithm in 1974, which, together with multi-layer perception-like neurons (e.g., Neocognitron) brought life to the ANN field again [79]. The backpropagation algorithm (abbreviation for 'backward propagation of errors') is used together with optimization methods such as gradient descent to updates the weights of the neurons in different layers based on the result of the loss function. The network is considered a supervised learning method because its measure of a successful classification brings the network closer to the desired output (Figure 3.1). However, an increase in the number of the network layers increases the problem of vanishing gradient.

Different frameworks have been tried since the 1980s, and some of these frameworks were few layers deep. However, the creation of deep networks has been tried by some researcher, giving rise to the field of deep learning which became popular in 2006 through the works of Geoff Hinton, Yann LeCun and other researchers on deep networks in machine learning [80,81].

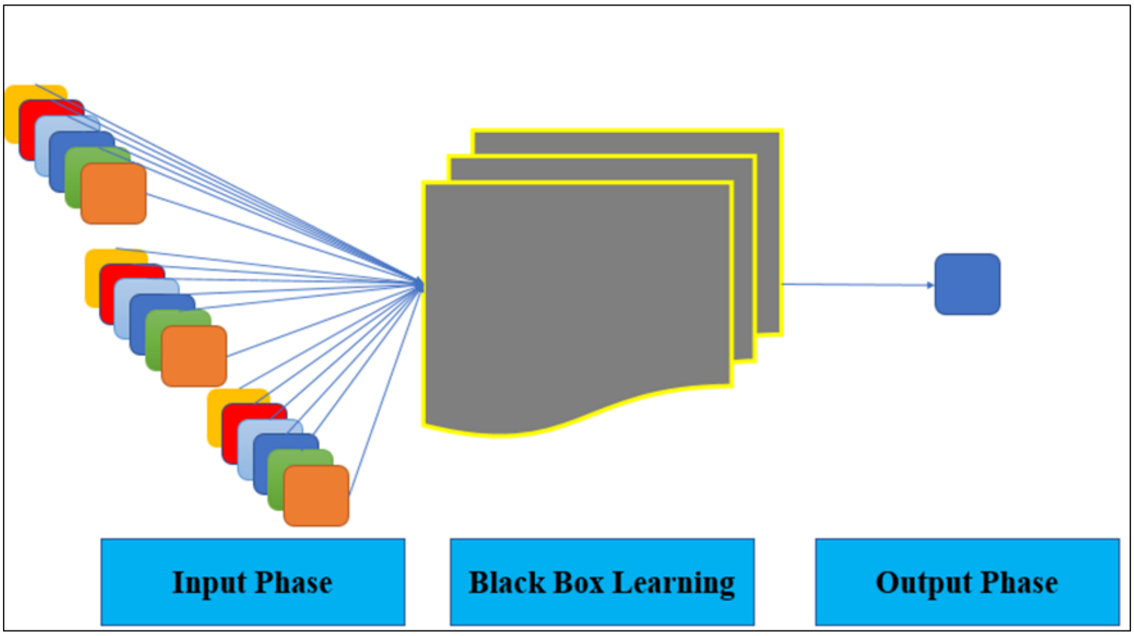


Figure 3.1 A neural network that may sit inside a neural box performing machine learning

3.2 Types of Machine Learning

Neural networks are mainly characterized by their ability to learn [82]. They can be trained by simply providing a series of example inputs and the expected outputs; pass them through a mechanical process to take the weights from the initial random values to produce more accurate predictions. However, this process cannot be performed in complex problems like image and speech recognition, where the required networks increase in size with the number of weights.

Being that deep learning is a subfield of machine learning, it is important to point out the existing types of learning. Machine learning algorithms are better classified as either supervised or unsupervised based on the intended goals and the previously available input data information [83].

3.2.1 Supervised Learning

This involves learning from a dataset that is previously labelled. Such dataset consists of the input object and its label/class. Supervised learning frameworks produce a function that maps the new examples using the training data (Figure 3.2) [84]. The algorithm can correctly classify unseen instances in the best-case scenario into their respective classes.

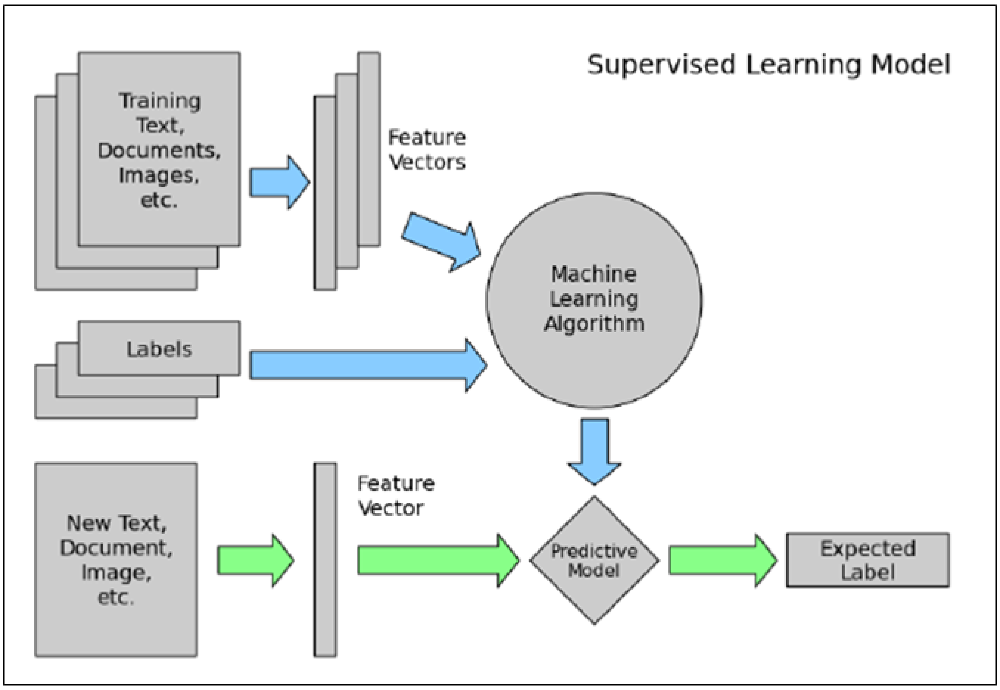


Figure 3.2 Supervised learning flow chart[65]

3.2.2 Unsupervised Learning

This involves finding the underlying structure of a previously unlabelled data. Cluster analysis is a common way of producing object clusters based on their similarity with the variables of other objects (Figure 3.3) [86].

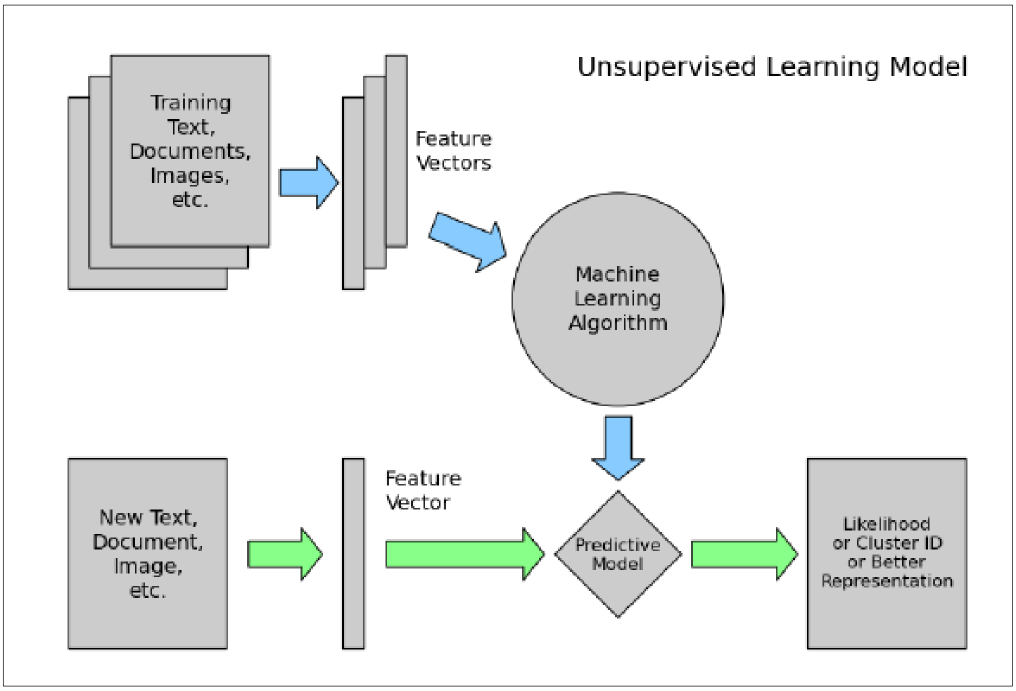


Figure 3.3 Unsupervised learning flow chart[66]

3.3 Overview of Deep Neural Network

As a branch of machine learning, deep learning is an assemblage of learning methods and algorithms which usually employ nonlinear units transformation units (such as Restricted Boltzmann Machines or artificial neurons) in a hierarchical layered manner [87]. With the aid of learning algorithms, different input data abstractions are learned from lower to higher abstraction; such abstractions are lower in layers near the input data but higher in layers further away.

The primary reason deep learning took off so quickly is that it offered better performance on many problems. But that's not the only reason. Deep learning also makes problem-solving much easier, because it completely automates what used to be the most crucial step in a machine-learning workflow: feature engineering. Previous machine learning models are shallow learning and only involved transforming the input data into one or two successive representation spaces, usually via simple transformations such as high-dimensional nonlinear projections such as SVR or decision trees. However, the refined representations required by complex problems generally can't be attained by such techniques. As such, humans had to go to great length to make the initial input data more amenable to processing by these methods: that is, they had to manually engineer good layers of representations for their data. This is called feature engineering. Deep learning, on the other hand, completely automates this step: with deep learning, you learn all features in one pass rather than having to engineer them yourself. This has greatly simplified machine-learning workflows, often replacing sophisticated multistage pipelines with a single, simple, end-to-end deep-learning model.

Studies in deep learning are often inspired by intuition, empirical results, current neuroscience knowledge, and theoretical arguments from circuit theory [35,87–89]. It is a progressing field of machine learning owing to the good application results it produces. Informally, it could be said that deep learning methods are used to extract some important information from a data structure instead of performing a shallow statistical inference on unstructured dataset such as matrices and vectors of values.

Deep learning differed from the other machine learning algorithms in its depth. The other machine learning methods which are not deep in terms of stacked layered units are Support Vector Machines, Naïve Bayes and Decision Trees [88]. These methods are often referred to as shallow architectures; they apply fewer input signals/data transformation as they migrate from the input layer to the output layer.

Several problems can be solved using the application of neural networks due to their ability to calculate any computable function. They are mainly useful in solving problems that can tolerate some levels of error or problems that are laden with several historical data but cannot be easily handled via the application of the hard and fast rules [90,91]. The study on the ANN over the past few decades formed the basis for the deep learning concept. Neural networks (NNs) are constructed from several layered interconnected nodes called neurons. In a typical feed-forward neural network, there is at least an input layer, a hidden layer, and an output layer. The number of features or attributes to be fed into a neural network corresponds to the number of nodes in the input layer and are analogous to the covariates or independent variables that will be incorporated in a linear regression model. The number of items is predated or classified is represented by the number of nodes in the output nodes. The nonlinear transformation of the original input attributes is performed using the hidden layer nodes.

The construction of a standard NNs requires the use of neurons to produce real-valued activations, and the NNs can behave as expected by adjusting the weights of the neurons. There may be several chains of computational stages during the training of an NNs depending on the problem to be solved. Since 1980, backpropagation, an efficient gradient descent algorithm, has played a significant role in NNs by its capability of training ANN via a teacher-based supervised learning method [92]. The performance of backpropagation during the testing of data is not usually satisfactory, although it presents a high training accuracy. One issue with backpropagation is that it is often trapped in local optimal because it is based on local gradient information with a random initial point. Furthermore, there is a problem of over-fitting if the training data is not reasonably large enough [93]. Based on these issues, several effective machine learning algorithms such as SVM, ANFIS, genetic programming which attain global optimum at lower power consumption have been used.

The layer-wise-greedy-learning method was proposed in 2006 by Hinton to mark the introduction of deep learning techniques [80]. This learning method was proposed based on the fact that a network should be pertained via an unsupervised learning process before being subsequently trained by the layer-by-layer training. The dimension of the data can be reduced by extracting features from the inputs to obtain a compact representation. The samples will then, be labelled by exporting the features to the next layer, and the labelled data will be deployed to fine-tune the network. There are two reasons attributable to the popularity of deep learning methods: (i) the issue of data overfitting can be addressed by the development of big data analysis techniques. (ii) non-random initial values will be assigned to the network during the pre-training procedure before the unsupervised learning. Therefore, a faster coverage rate and a better local minimum can be achieved after the training process.

Though several types of deep learning model exist, the focus of this discussion is on the deep neural networks that are constructed from multiple hidden layers, often known as backpropagation neural networks [80]. Deep learning is historically based on how to use backpropagation with gradient descent and a large number of nodes and hidden layers. This type of backpropagation neural network is indeed the first deep learning approach that showed a wide range of application. A typical deep neural network (DNN) comprised of closely embedded input, output, and several hidden layers. The input and hidden layers are directly connected and function together to weigh the input values to produce a new set of real numbers that will be transmitted to the output layer (Figure 3.4). Finally, the output layer, based on the transmitted values, classify or predict the outcome of the process.

The main merit of the DNN is that the deep multi-layer neural network is made up of several levels of nonlinearities which made them applicable in the representation of highly nonlinear and/or highly-varying functions [85,94]. They can identify complicated patterns in data and can be applied in complex natural problems. The connection weights connections between the layers, as in the single-layer neural network, are updated to ensure the closeness of the output value to the targeted output.

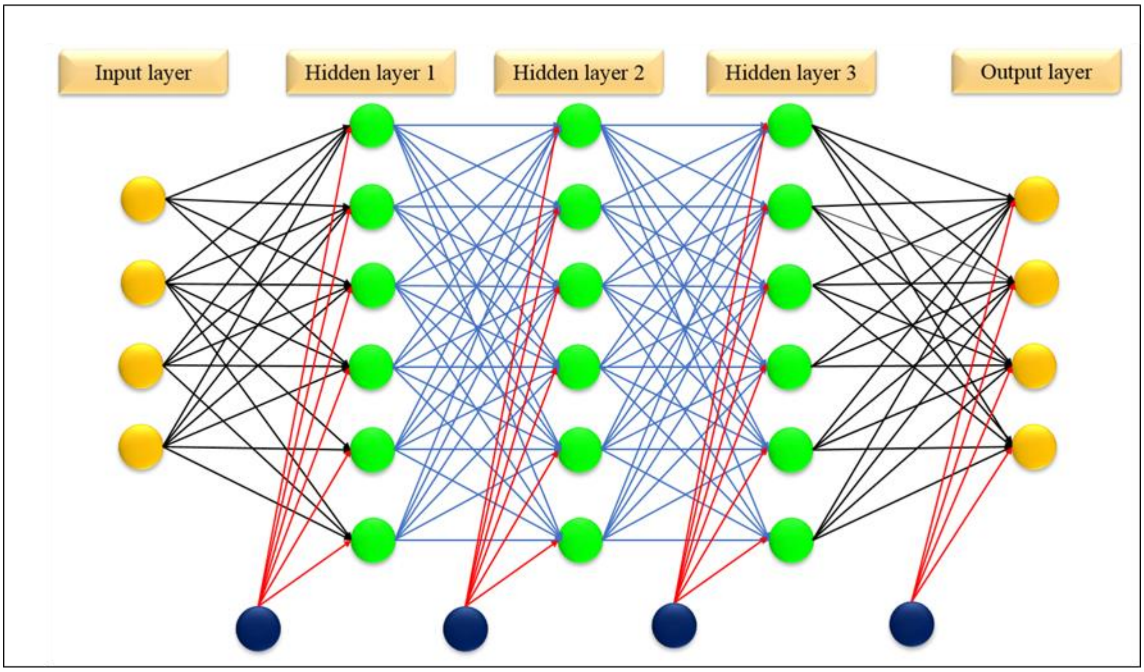


Figure 3.4 The standard architecture of deep neural network description [85].

Figure 3.4 describes the general architecture of the DNN predictive model with hidden layer, "and for the current research for sıngle DNN were 3 layers were set as per the RapidMiner software default". Input variables layer denotes as (0 layers) and (L layer) represents the output variable layer. The mathematical procedure can be described as follows [95]:

$$M^l = f(w^l \varphi^{l-1} + b^l) \text{ for } 0< l < L \tag{3.1}$$

where f is the activation function, w is the weight matrix, b is the bias. In this study, the implemented activation function for the excitation vector is sigmoid function owing to its applicability for a regression problem.

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{3.2}$$

Note that the outcome of $\sigma(z)$ is limited between (0-1), that emphases the sparse. However, it is a systematic activation function. The DNN model was developed using Rapidminer software. A thematic snap from the modelling software is reported in Figure 3.5.

3.4 Support Vector Regression Model

In 1995, Vapnik proposed the support vector machine (SVM) as an optimization method which tries to separate a given training set by establishing a hyperplane within the original input space and allowing enough distance from the nearest instances on both sides to the hyperplane [96]. In a regression problem, the SVR model approximates the error between the input and output variables [97]. The errors are equal to the limited marginal of the SVR learning range as denoted in Figure 3.6. The investigated problem in this research is featured by nonlinearity pattern in which the mapping of the SVR model characterized by high dimensional space, also known as feature space. The notation presentation of the SVR model can be expressed as followed. Assuming there is given a set of training data set represented by M:

$$M = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \tag{3.3}$$

where x and y the input and output information. The regression nonlinear function implemented here to be solved [98]:

$$f(x) = w^M \varphi(x_1) + b \tag{3.4}$$

Note that the w denotes the weight vector, $\varphi(x_1)$ presents the high order of the feature space; whereas, the last variable of the function is the bias (b). Well, the main goal of this regression function is to determine the output based on the training data set M, with a certain deviation of error called loss function ε from the actual observation of the whole training data set. Hence, this can be described through the constrained convex optimization function [98]:

$$Minimum\ \varphi(w, \xi) = 0.5\, \|w\|^2 + C\left(\sum_{i=1}^{l} \xi_i\right)\varphi(x_1) + b \tag{3.5}$$

$$\text{Subjected to: } M_i(w^M x_i + b) \geq 1 - \xi_i, \xi \geq 0 \tag{3.6}$$

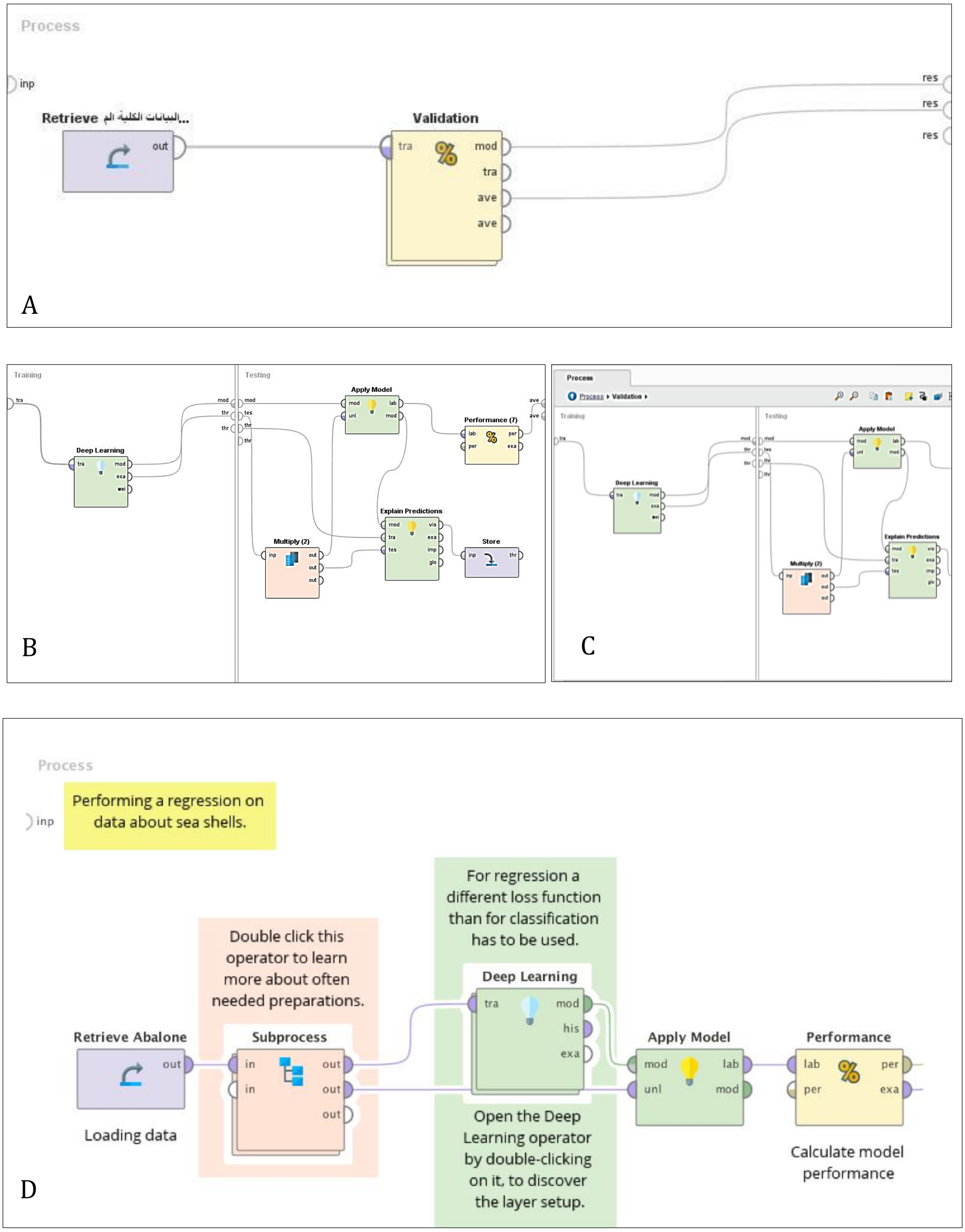


Figure 3.5 (A, B, C & D) A thematic map for the DNN model using Rapidminer software.

The definitions of the convex formulation are (ξ: slack variable) and (C: the positive regularization). Note that ξ penalizes the training error through the loss function for the selected error tolerance. Whereas, the positive parameter shrinks the weight variables during the optimization process. Figure 3.6 reports the structure of the SVR model.

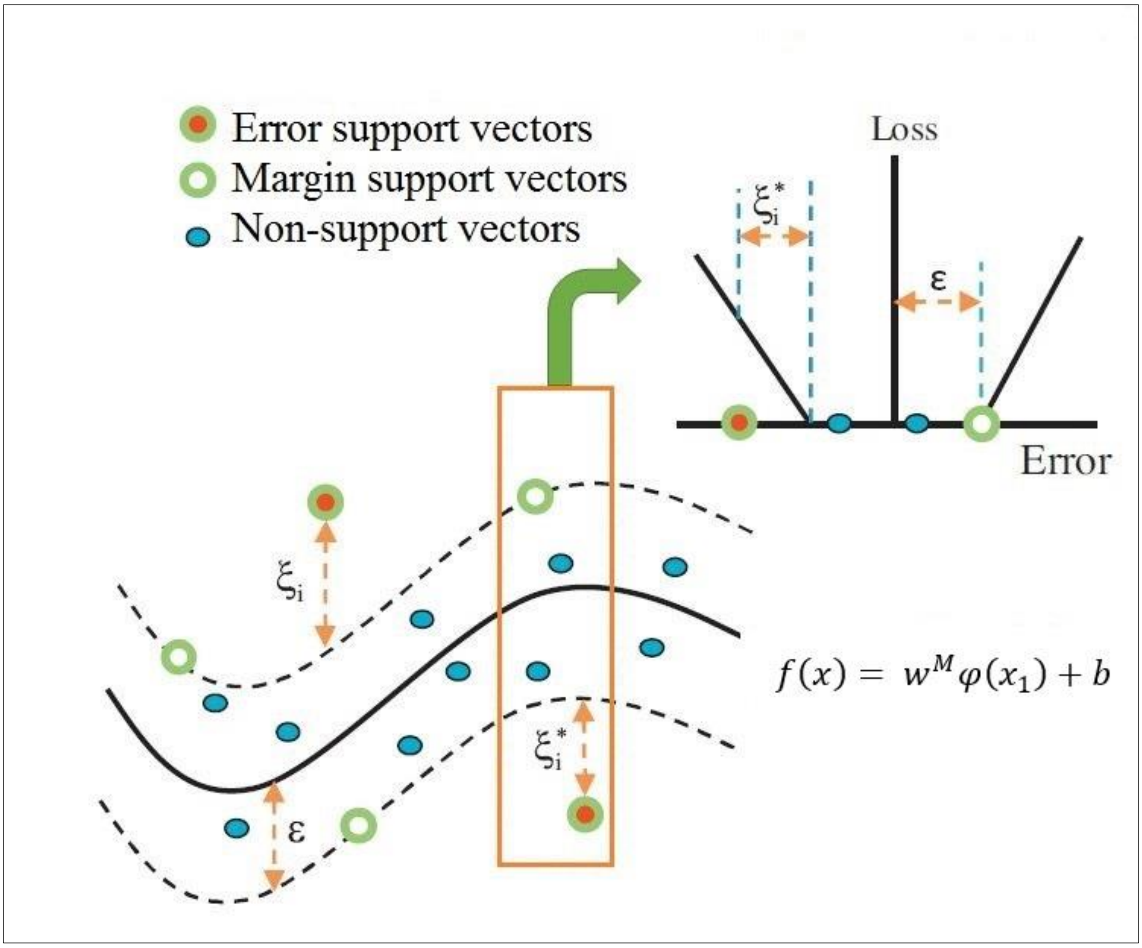


Figure 3.6 The structure of the support vector regression model [99].

The optimization problem of the SVR model usually elucidated using the Lagrangian multipliers, minimal sequential optimization [100]. The radial basis kernel function is employed for the feature mapping of the training data sets. The internal parameters of the radial basis function tuned using the grid-search approach. SVR model is developed as a comparable model using the Rapidminer software as presented in Figure 3.7.

3.5 Genetic Algorithm (GA) Optimization

GA is a very well-known optimization technique that can be classified as an evolutionary method based on biological process [42]. The effectiveness of this optimization approach discussed comprehensibly in term of solving the nonlinearity and stochasticity by [43]. The main processes are involved in the implementation of the heuristic GA are including reproduction of chromosomes, crossover, and mutation. Note these processes are applied to satisfied the probability of the discretization of the input variables that are coded into binary strings [44-46].

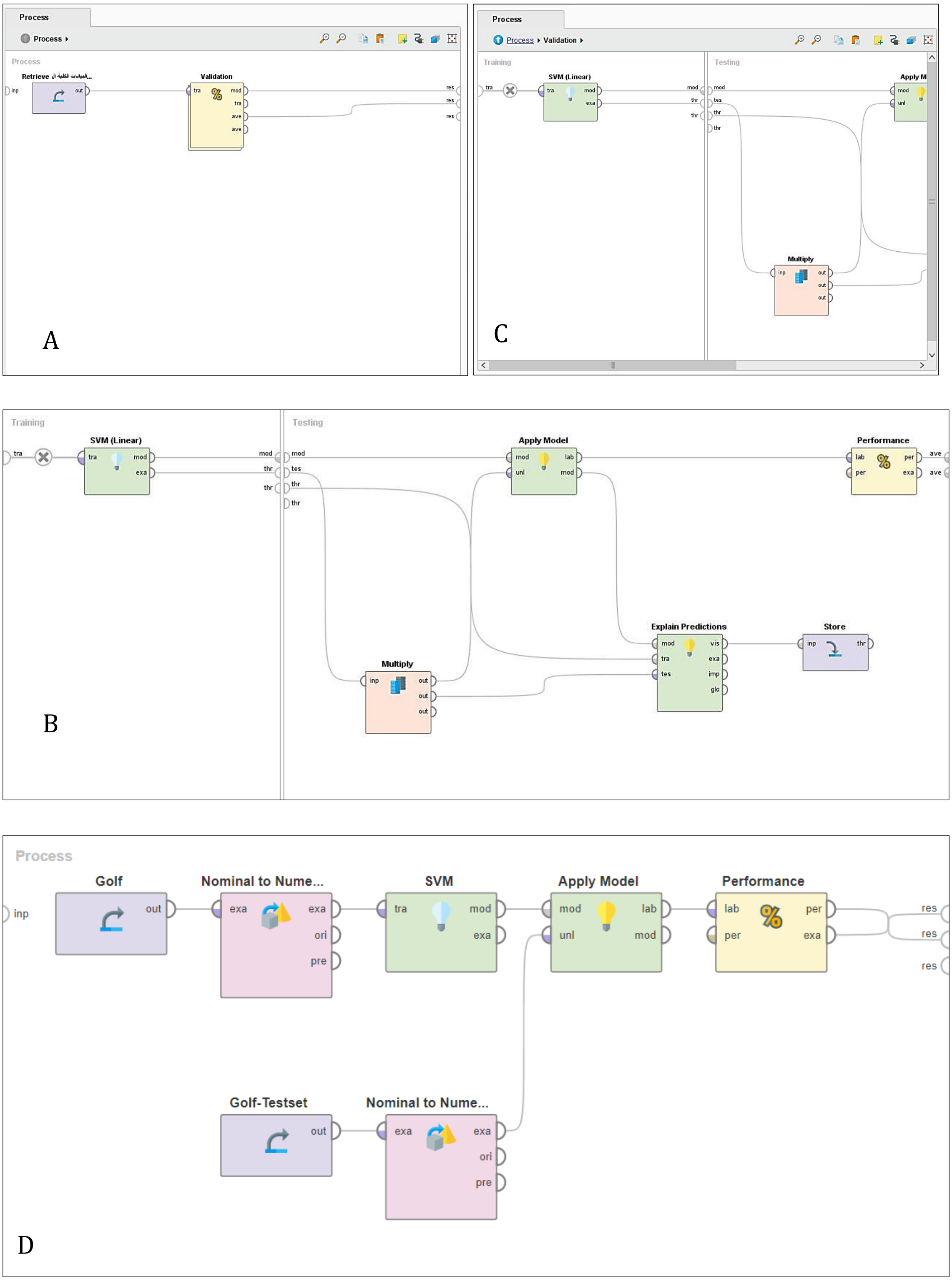


Figure 3.7 (A, B, C & D) A thematic map for the SVR model using Rapidminer software.

The GA processes, integrated with the DNN predictive model, are presented in Figure 3.8. which clearly shows that prediction process consists of two parts the first one Genetic algorithm phase (GA) is a branch of AI and evolutionary algorithms which is one of the modern approaches of numerical optimization that is based on Charles Darwin's theory of "survival of the fittest" and "natural selection". GA, a search algorithm based on concepts of natural selection and genetics, was officially introduced by Holland in the 1990s [101]. The underlying principles of GA are to generate an initial population of chromosomes (search solutions) and then use selection and recombination of operators generate a new, more effective population which eventually will have the fittest chromosome (optimal value) among them.

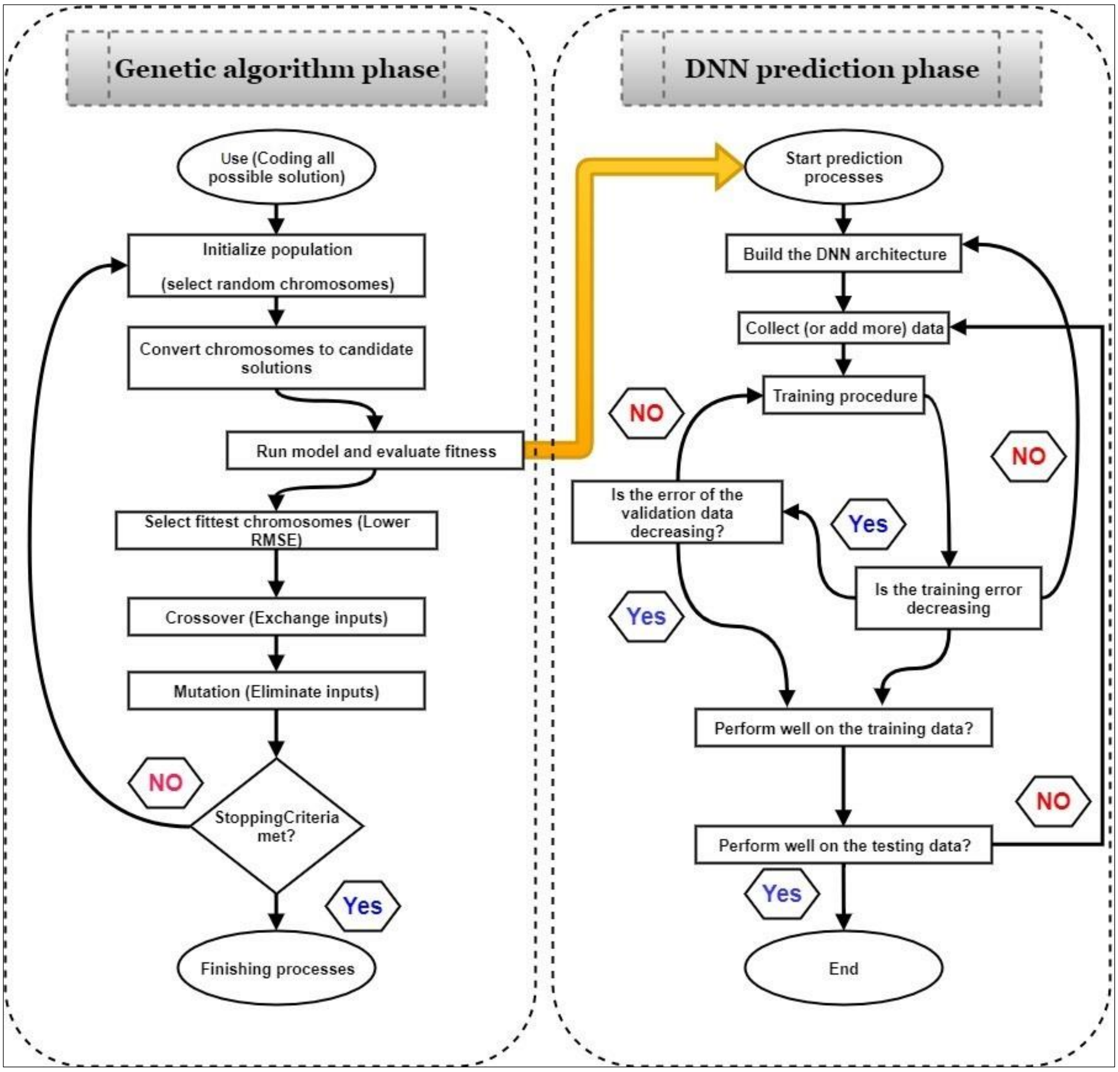


Figure 3.8 The proposed hybrid genetic algorithm deep neural network (GA-DNN) predictive model.

3.5.1 Hybrid Predictive Model Creating

The ten steps of operation of GA and DNN hybrid intelligence are as follows.

Step 1: Use

This step is including Coding all possible solutions.

Step 2 (initialization of population).

Generate an initial population of chromosomes which are bit strings of randomly generated binary values.

Step 3 Converting.

In this step convert chromosomes to candidate solution.

Step 4 (Run model and fitness evaluation).

Take the prediction accuracy of each chromosome from GA as its fitness value for DNN.

Step 5 (selection).

Select chromosomes were to cross over using tournament selection technique. A tournament selection involves running several tournaments on a few chromosomes chosen at random from the population. The winner of each tournament is selected for crossover.

Step 6 (crossover).

Apply an arithmetic crossover operator that defines a linear combination of two chromosomes.

Step 7 (mutation).

Inject new genes into the population with uniform mutation operator and generate a random slot number of the crossed-over chromosome as well as flip the binary value in that slot

Step 8 (stopping criterion).

Determine whether to continue or exit the loop. The stopping criterion was repeated until finishing all the selected data.

Step 9 (replacement).

Replace old chromosomes with two best offspring chromosomes for the next generation.

Step 10 (loop).

Go to Step 2 to repeat the process.

The connection between GA and DNN was allocated at step 4; the output of GA was taken into DNN prediction phase to start the prediction process as a hybrid model. Then build DNN architecture are created, which followed by collecting or adding more data. The next step is Training procedure finished by a conditional statement asking about the decreasing in training error (**is the training error decreasing?**) This step leads to three actions depending on the answer of this statement, the first one if the error increased (make a loop and go to **build DNN architecture** action**).** The second one if the error decreased (make a loop and go to **is the error of the validation data decreasing action**), the third actions if training error is still the same then go to **Perform well on the training data. (Is the error of the validation data decreasing action**), is a conditional statement and has two actions the first one if the error increased (make a loop and go to **Training procedure** action), the second one if the error decreased (then go to next step **Perform well on the training data).** The step of **(Perform well on the training data?)** is leading to **(Perform well on the testing data)** step directly, which is a conditional statement has two actions. The first one if the **testing data is not well Performed** (make a loop and go to **collecting or adding more data** action), the second one if the **testing data is well Performed** (go to **End**) and finish the process.

An evolutionary algorithm is an optimization approach that mimics the concept of natural evolution [102]. In the evolutionary algorithms, three basic concepts are involved: firstly, the parents create the offspring via crossover; second, the individuals within a generation have the chance of undergoing mutation (changes); and finally, the fitter individuals have a higher chance of survival (natural selection).

It is now certain that attribute subsets can be represented with bit vectors. Thus, there is a possibility of selecting all the features of a data set with ten features such as (1 1 1 1 1 1 1 1 1 1). The third attribute of the data set can be represented using a bit vector in the form of (0 0 1 0 0 0 0 0 0 0).

In the evolutionary algorithm, the first step is the creation of a population of individuals which evolves within the time [103]. This initial step is known as the initialization phase of the GA. In the starting population, the individuals are randomly generated and represented as a bit vector-like earlier described. These individuals can be created by tossing a coin for any available attribute. And based on the probability toss outcome, the attribute to be included in the population can be determined. No rules are governing the size of the initial population. However, there must be at least two individuals in a GA population to proceed to the crossover phase [104]. A perfect rule of thumb is the acceptance ranging from 5 to 30% of the total number of attributes as the size of the initial population. They have created the initial population; several steps need to be performed to reach the stopping criterion. A thematic map for the generated GA optimizer is reported in Figure 3.9.

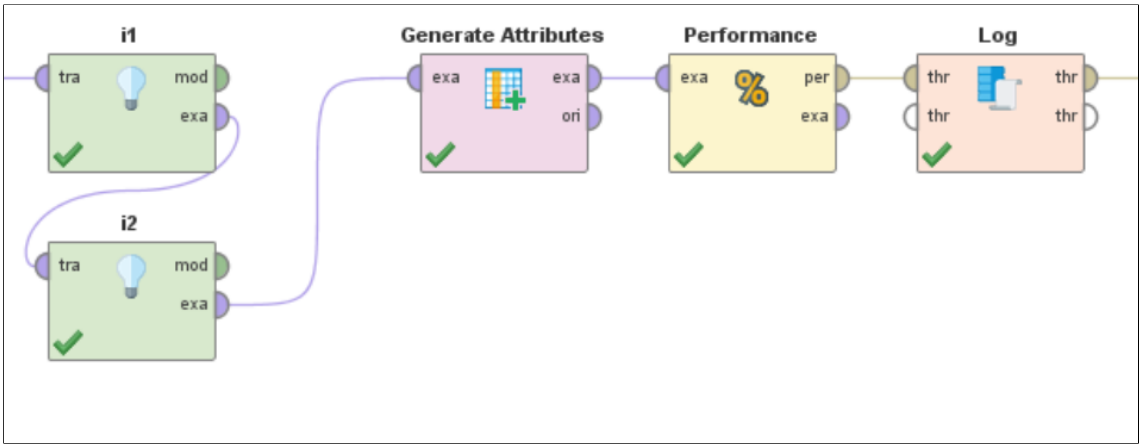


Figure 3.9 A thematic map for the GA model using Rapidminer software.

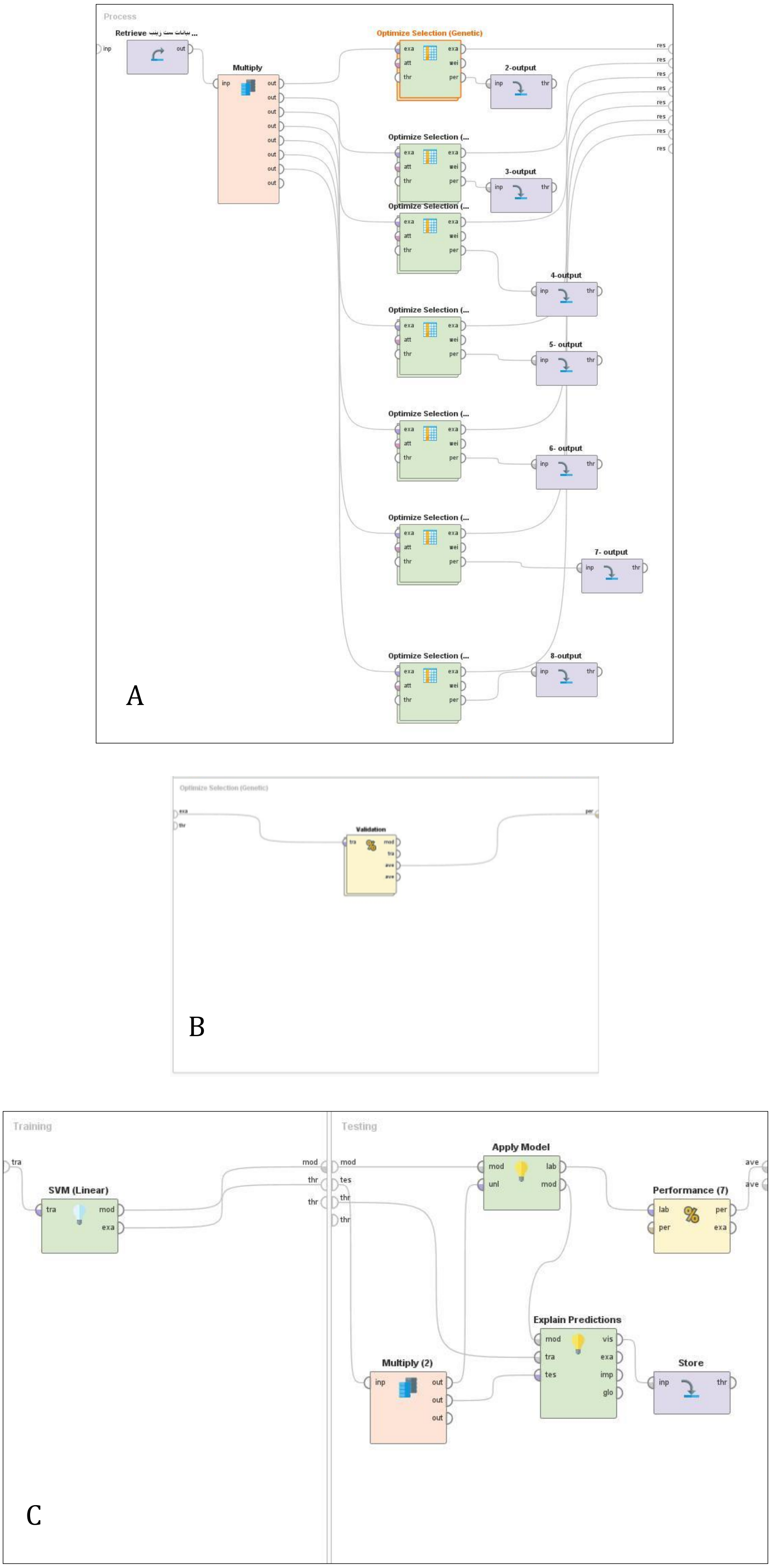


Figure 3.10 (A, B &C) A thematic map for the GA-SVR model using Rapidminer software.

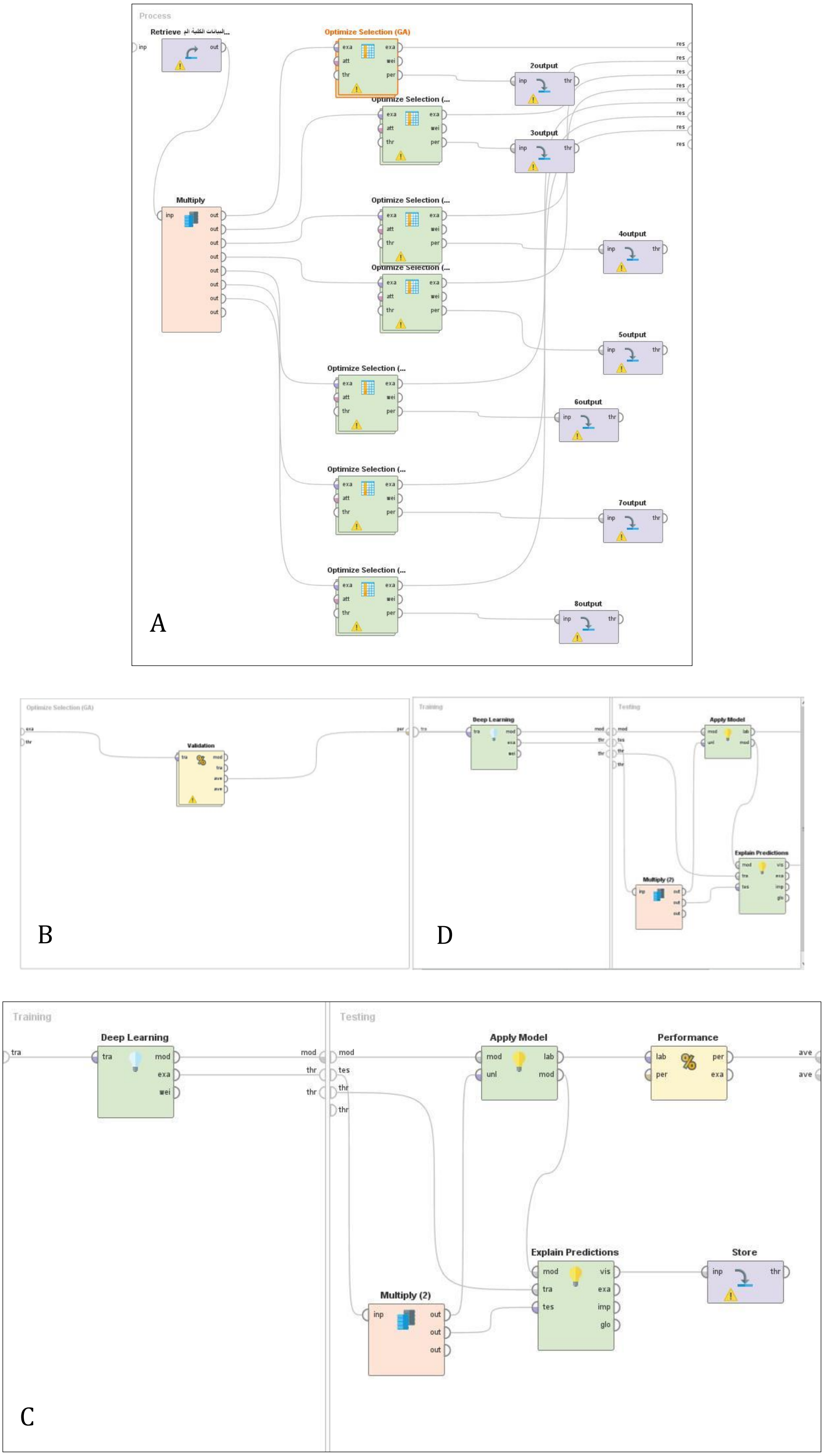


Figure 3.11 (A, B, C & D) A thematic map for the GA-DNN model using Rapidminer software.

3.6 Brute-force Input Selection

Brute-force (BF) is a systematic selecting approach that solving problems which require the enumeration of all the possible features [105]. This is sake of achieving a solution to a specific problem and checking the suitability of each option towards satisfying the problem statement [106]. BF is usually performing to find the divisors of a number n would list all the integers from 1 to n and check that each integer will perfectly divide n without any remainder. Although a BF search is easy to implement and will always establish a solution to the problem, its cost is directly related to the number of options considered, and this number tends to grow with the size of the problem in many practical situations. BF is therefore applicable in situations where the size of the problem is limited or in the absence of a specific heuristic method that can be used effectively to reduce the number of solutions to a considerable size. BF approach can also be used as a yardstick for benchmarking the performance of other algorithms. It is considered as one of the simplest searching approaches. Its potential inspired the selection of this searching approach to be integrated with the developed predictive model in the feature selection problem.

3.7 Modelling Development and Prediction Skills Metrics

The current research is conducted on fifteen construction projects executed in Baghdad city, Iraq. The researcher has conducted a field study of a group of construction projects located in Iraq. Fifteen completed construction projects in which delays were observed, information on each project shown in Table 3.1, were used for researching the problems of the delays and identifying the problems that occurred and led to project delay. It was obtained information about the delay in the completed projects under study by reviewing the records and documents of each project (such as the contract document, variation orders, additional durations, payments, and design drawings). Also, other required information was collected from the local authority in Baghdad such as (The Ministry of Construction, Housing, Municipalities, and Public Works in Iraq and the secretariat of the capital, Baghdad, and the offices followed the capital's secretariat in addition to private companies

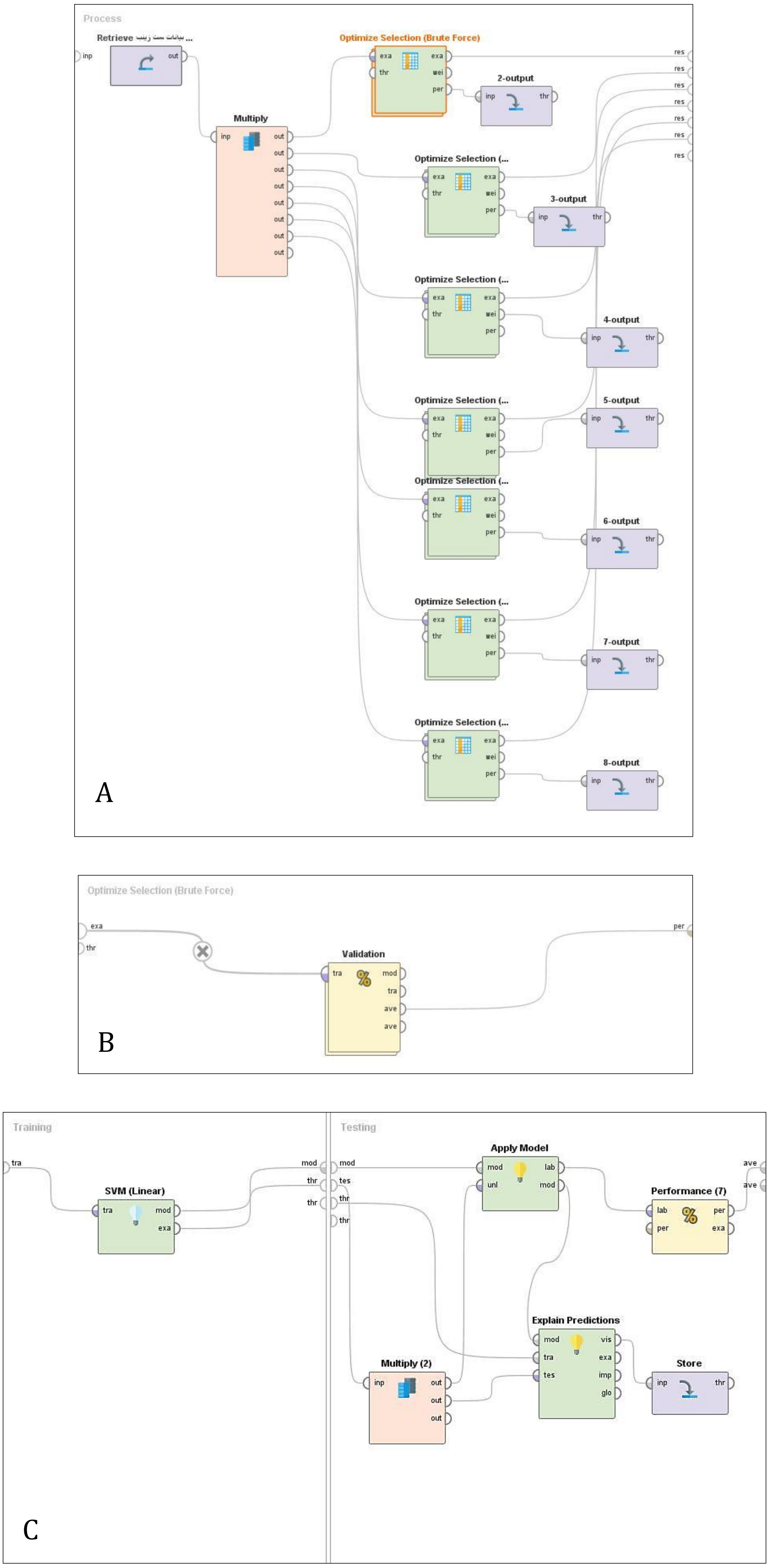


Figure 3.12 (A, B, & C) A thematic map for the BF-SVR model using Rapidminer software.

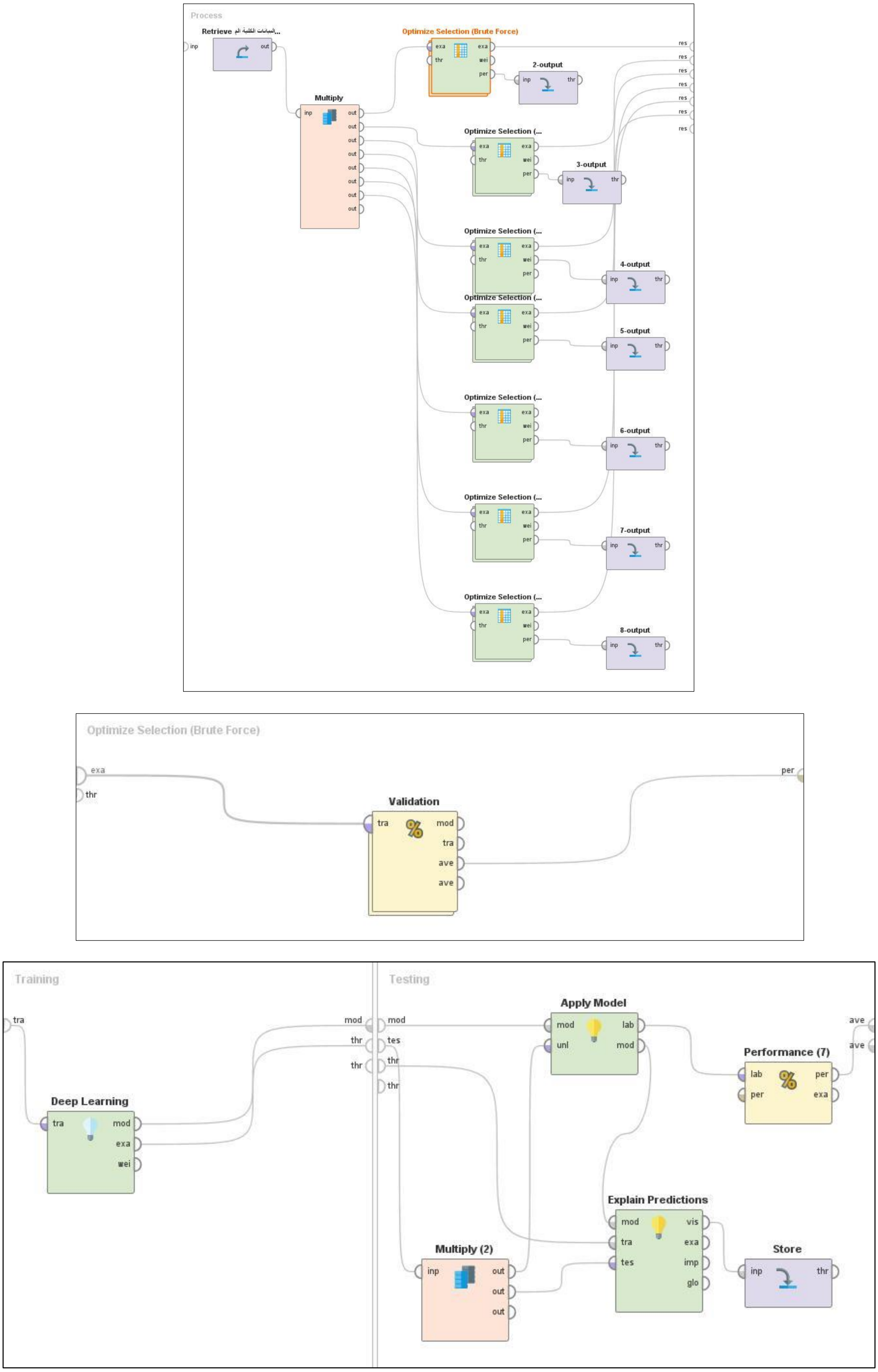


Figure 3.13 (A, B, & C) A thematic map for the BF-DNN model using Rapidminer software.

such as ALBILAL GROUP-General Contracts Co.Ltd and Al-Yamamah Engineering Company). Some of the available information was collected from the website of the ministry

(https://www.moch.gov.iq/%D9%85%D8%B4%D8%A7%D8%B1%D9%8A%D8%B9%20%D8%A7%D9%84%D9%88%D8%B2%D8%A7%D8%B1%D8%A9%20-%20%D9%85%D8%A8%D8%A7%D9%86%D9%8A) such as the projects finishing and starting date, while the other information presented personally provided for the researchers, academics and higher education students (this action well known in Iraq). The details information about those projects is provided in Table 3.1. The construction duration of the projects is ranged between nine to fourteen months. Established construction projects are related to residential projects.

Table 3.1 The biodata of the inspected construction projections

Project name	**Total area (m2)**	**Under ground floors**	**Ground floors**	**Buildings**	**Start date**	**Finish date**	**Duration (days)**	**Contract amount ($)**	**Prediction periods**
1	7,854	2	2	2	4/22/2007	4/17/2008	361	4,319,000	9
2	6,238	1	1	1	01/03/2008	10/24/2009	295	3,512,900	12
3	7,284	0	1	1	09/28/2006	08/31/2007	337	5,119,050	10
4	6,824	0	2	1	02/15/2005	01/17/2006	336	4,519,050	13
5	6,453	1	1	2	06/05/2005	07/02/2006	392	4,812,800	10
6	7,471	2	1	2	10/05/2008	08/12/2009	312	3,627,300	12
7	7,864	1	1	1	03/03/2009	01/27/2010	330	4,423,050	11
8	6,678	1	1	1	05/01/2007	02/24/2008	299	3,627,300	9
9	9,340	1	2	3	06/23/2007	08/13/2008	417	6,339,350	11
10	10,628	0	2	2	10/09/2004	11/29/2005	416	6,128,150	13
11	8,245	0	1	3	08/15/2010	07/13/2011	332	5,111,350	11
12	8,782	0	2	3	01/10//2006	01/05/2007	360	3,914,600	12
13	6,625	1	2	1	01/01/2008	02/28/2009	424	7,529,350	14
14	5,441	1	1	2	08/25/2007	08/18/2008	359	5,223,100	13
15	5,730	1	1	1	04/01/2005	02/27/2006	332	5,522,500	14
Total									174
Training									131
Testing									43

In this research, a free data mining software called Rapidminer software [107]. Overall, RapidMiner was allowing rapidly trying out different machine learning models and comparing each result with one another. It also allowed to conveniently addressing my workflow without having to write code [108]. It is a great tool for students and people without a strong programming background. Its well-documented functions. RapidMiner has many advantages such as devoiding of one of the daunting requirements for data scientists and students is learning a programming language such as Matlab and python and writing code for their tasks. This is on top of having to analyze and learn the complex algorithms needed for the task. This can be a time-consuming problem, especially for those who are not adept at programming. However, this issue is left in the past after appearing RapidMiner Studio, where RapidMiner features are drag and drop visual interface, which makes all the difference. Data preparation to the final output and visualization is as simple as dragging blocks of your workflow into a canvas and connecting them. RapidMiner Studio also has most of the machine learning models used in the academe and the industry [107–109]. One of the difficulties when dealing with code is tweaking the parameters of these models, but the visual interface, you could simply click on the process and update this. Each of the processes has its description, input, output, and parameters well described. Tutorial videos, as well as blogs, are available on their website for learning this software easily. Finally, RapidMiner Studio has a community of data scientists that can help you when you have a question [110].

The collected information of the projects is including cost variance (CV), schedule variance (SV), cost performance index (CPI), schedule performance index (SPI), subcontractor billed index, owner billed index, Climate effect index, change order index, construction price fluctuation (CCI). Whereas, the estimate at completion is the main targeted variable to be estimated. The nine factors are used as predictors to determine the EAC. The 15 projects comprised 174 periods, 75% of the total periods (131 periods) is performed for the training phase and 25% (43 periods) for the testing phase of the predictive models. The data size was varied from project to another one, depending on the data availability, the prediction model type and the effective factors on the prediction of EAC. For example, Bayram [111] used (530) training data for (Construction cost, gross floor area, and building height) to estimate construction time. Kim et al., [112] used (47) training data (Gross floor area, no. of stories above ground, no. of stories below ground, structure type, whether the construction is undertaken in winter, and building area) for prediction of EAC. Kaka and Price [113] used (801) training data (Construction cost, client (public/private), construction type (building/civil-engineering), tender method (open/private/negotiation), and contractual arrangement (fixed-cost, cost-plus)). According to [114] trained a 4-layer DNN for the spectral enhancement. Three consecutive spectral envelopes were used as the input and output of the DNN. The FFT length for calculating spectral envelopes was set to 4096, which leads to 3 × 2049 = 6147 units in both input and output layers. The number of units in each of the two hidden layers was set to 2048. RBMs and BBAMs were trained using the contrastive divergence (CD) algorithm. The DNN was trained using paired synthetic and natural spectra aligned using dynamic time warping (DTW).

Sivapatham, et al., [115] they using 256 data entry and applying GA-DNN integration to enhance the quality and intelligibility of the noisy speech. In this proposed model, the Voiced Speech (VS) T-F mask is computed using correlogram, frame energy and cross-channel correlogram and Unvoiced Speech (UVS) T-F mask is computed using speech onset/offset. The T-F mask obtained using speech onset and offset represents both voiced and unvoiced segment of the noisy speech signal. The results of the proposed model shows a prompt improvement in the speech quality and intelligibility with average of 0.73, 4.07, 0.17, 0.26 and 0.22 for PESQ, SNR, STOI, CSII and NCM when compared with the existing speech separation systems.

Cheng, et al., [2] studies the estimate at Completion for construction projects using Evolutionary Support Vector Machine Inference Model. And the selected parameters that used during prediction were Change order index, Schedule performance index, Subcontractor billed index, Subcontractor billed index, Climate effect index, Construction cost index, Change order index. On the other hand, the 269 periods collected from 11 cases were input into the database. An estimated value of prospective cost percentage for each case could be obtained via the ESIM performance assessment module. The estimated value is termed ‘Predicted output’.

Cheng, et al., [116] studies a novel time-depended evolutionary fuzzy SVM inference model for estimating construction project at completion. And the selected parameters that used during prediction were Construction progress (%), ACp

(Actual Cost percentage), EVp (Earned Value percentage), CPI (Cost Performance Index), SPI (Schedule Performance Index), Subcontractor Billed Index, Owner Billed Index Owner, Change Order Index and CCI (Construction Cost Index). On the other hand, The training and testing data sets consist of 269 and 37 data cases, respectively. Table 3 shows the 10 input variables from project C, which had 20 completion periods.

Nassar and AbouRizk [117] studies the Practical Application for Integrated Performance Measurement of Construction Projects. The priorities for the performance indexes presented in this study are derived according to importance using Saaty's analytical hierarchy process (AHP). And the selected parameters that used during prediction were Cost performance index (CPI), Schedule performance index (SPI), Billing performance index (BPI), Profitability performance index (PPI), Safety performance index (SFI), Quality performance index (QPI), Team satisfaction index (TSI) and Client satisfaction index. On the other hand, the proposed eight performance indicators include objective and subjective factors and are based on feedback and discussions with 15 major construction contractors representing a wide range of sectors: power, oil and gas, petrochemical, pipelines, residential, transportation, and utilities.

Barraza et al. [118] Probabilistic Forecasting of Project Performance Using Stochastic S Curves. The concept of stochastic S curves (SS curves) to determine forecasted project estimates as an alternative to using deterministic S curves and traditional forecasting methods. A simulation approach is used for generating the stochastic S curves, and it is based on the defined variability in duration and cost of the individual activities within the process. And the selected parameters that used during prediction were Final project performance is determined by comparing the planned budget and project duration, with the expected forecasted final cost and elapsed time, respectively. On the other hand, running 500 project simulations in SPECIESS, the expected estimations of project duration and final cost were obtained from the simulated SS curve values and from the correspondent cumulative distribution functions at project completion~100% of work complete.

Juszczyk [119] studied the Challenges of Nonparametric Cost Estimation of Construction Works With the Use of Artificial Intelligence Tools. Nonparametric cost estimation in construction projects with the use of artificial networks is presented as suitable mainly for the early estimates. These conceptual estimates are based on the variables – namely, cost predictors that characterize the project or a facility. Data gathered on the basis of completed projects are combined together and applied to the current project cost estimation process. The proposed approach is based on the concept of nonparametric cost estimation and application of artificial neural networks.

One obstacle to realizing smart applications is the large amount of data volume of video streaming. For example, Google's self-driving car can generate up to 750 megabytes of sensor data per second [120], but the average uplink rate of 4G, fastest existing solution, is only 5.85Mbps [121]. The data rate is substantially decreased when the user is fast moving or the network is heavily loaded. In order to avoid the effect of network and put the computing at the proximity of data source, edge computing emerges. As a network-free approach, it provides anywhere and anytime available computing resources For example, Deep neural networks (DNNs) to analyze more that 2000 visual imagery [122].

Deep neural networks (DNN) have achieved breakthroughs in applications with large sample size. However, when facing high dimension, low sample size (HDLSS) data, such as the phenotype prediction problem using genetic data in bioinformatics, DNN suffers from overfitting and high-variance gradients. In this research, they propose a DNN model tailored for the HDLSS data, named Deep Neural Pursuit (DNP). DNP selects a subset of high dimensional features for the alleviation of overfitting and takes the average over multiple dropouts to calculate gradients with low variance. As the first DNN method applied on the HDLSS data, DNP enjoys the advantages of the high nonlinearity, the robustness to high dimensionality, the capability of learning from a small number of samples, the stability in feature selection, and the end-to-end training where they use sample number ranging from (62-187) data. We demonstrate these advantages of DNP via empirical results on both synthetic and real-world biological datasets [123].

An excellent example of the successes in deep learning can be illustrated with the ImageNet Challenge [124]. This challenge is a contest involving several different components. One of the components is an image classification task where algorithms are given an image and they must identify what is in the image. The training set consists of 1.2 million images, each of which is labeled with one of 1000 object categories that the image contains. For the evaluation phase, the algorithm must accurately identify objects in a test set of images, which it hasn't previously seen. Depending on the previous studies, therefore the selected data (174) can be considered as appropriate for prediction of EAC.

A simple structure for the proposed predictive hybrid model exemplified in Figure 3.14

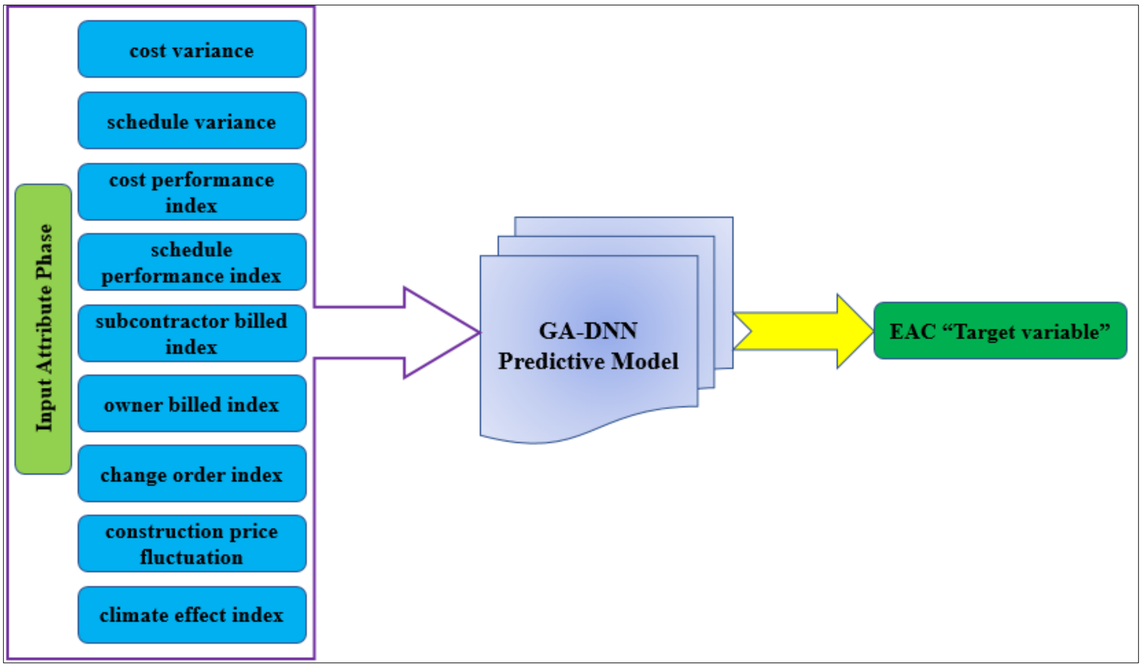


Figure 3.14 Input-output variables system structure using the hybrid intelligent GA-DNN predictive mode

CV is a quality indicator since it tests the performance (or failure) of combining scope, expense and timeline. It also suggests progress (or failures) in organizational training, procurement, and communications. The optimal utilization of energy (for example getting the correct thing correctly, the first time) is a significant advantage of efficiency [125]. PMI has reported a resource utilization measure that is derived from differences from real costs and budgeted costs — typically named cost variance (CV). This definition is expressed by the equation (CV = BCDP-ACDP), at which cost variance is proportional to the budgeted expense of the produced deliverables minus the real cost of the generated deliverables. Also, BCDP is recognized as earned value (EV) and reflects the accumulated equity of the generated deliverables [126]. The variance of Cost is the disparity in expense between the budgeted sum and the sum spent on an actual deliverable or a collection of deliverables. CV is not just an indication of resource utilization but also a metric of performance or loss in the combination of distance, expense and timing. Low or no CV for specific activities (for example, fewer than a few percent) usually implies that procedures, projects, and their systems are handled efficiently, and that procurement, communications, and human resources are adequately controlled [127]. CV is often an accurate indicator of a variety of non-quality characteristics. Broad CVs (for example, double-digit percentages) suggest that it lacks at least one of the consistency characteristics. Hence, large underruns in construction costs can even signify consistency issues.

Departures from expected timelines often show issues with the efficiency. The widely applied timeliness measure is schedule variance (SV) and that is the variance between the budgeted cost of the deliverables delivered and the project budget of the scheduled deliverables. The equation is expressed as: SV = BCDP-BCDS or SV = EV-BCDS [128,129]. SV calculates in cost units (for example, dollars) the discrepancy between the budget for the generated deliverables (BCDP) and the budget for the planned deliverables (BCDS) at the scheduled date selected. Therefore, the SV (as of a selected date) is the value obtained, minus the value expected to be received (BCDS) by that time. A positive SV means that more work was performed than had been expected by the time, while a negative SV indicates less research was accomplished [130].

The cost performance index (CPI) is the average of the benefit received (BCDP) to the total expense (ACDP). If the CPI is higher than 1, this indicates that the completed deliverables cost less than budgeted. If the CPI is lower than 1.0, the actual costs are higher than the costs budgeted. The CPI represents an effective measure to help project managers decide which project managers and projects require their attention. For example, if there are three projects (A, B and C) with CPI = 1.20, 0.75 and 1.05, respectively. Project (B) needs attention, whereas, Project C with a CPI of 1.05 is probably OK. Project (A) probably needs the director’s attention. Its CPI of 1.20 can indicate quality issues linked to big underruns. Instances of these quality issues are mentioned above (see "Cost Variances Measure Quality" section). Project managers for multiple quality-problem programs require the guidance of the manager [42,131–133].

The widely utilized measure of schedule variation is the Schedule Performance Index (SPI), that is calculated as the quotient of the cumulative budgeted cost of the produced deliverables, divided by the cumulative budgeted cost of the expected deliverables. Expressed more generally, the SPI is the ratio of the generated deliverables to the planned deliverables. Project management practitioners recognize fluctuations in expenses as a widely employed success metric. The test has not interacted with other behaviour. Schedule Performance Index (SPI) is the earned-value ratio (BCDP) to the scheduled-day value (BCDS). As SPI reaches 1.0, that implies more deliverables were generated than expected. When SPI is less than 1.0, deliverable production is behind schedule. The Project Efficiency Index (SPI) was used for calculating changes in plan schedules. SPI is determined as the ratio of the Deliverables Budget generated to the Scheduled Deliverables Budget. When SPI is 1.0, the delivering are on time; less than 1.0, off-plan; and more than 1.0, sooner than expected [129,134–136]

Climate change increases the risks of beginning work on a period of building projects [137]. Climate change can actively influence the building industry by environment and atmosphere, but it can also create indirect impacts such as site planning, disruptions, additional costs, the health of staff, material costs and delivery [138]. There's ample proof that risks were generated by climate change. Project management professionals ought to adapt their skills and knowledge to help appreciate the impact that climate change could have on the way they protect their potential ventures. They will know how to reduce the effect of climate change on their career, and continue to thrive and expand as production companies whilst also fulfilling their clients' needs [139].

The construction cost Index (CCI) is a weighted overall price index of constant material quantities[140,141]. This index offers short- and long-term adjustments in prices in an effort to gain more reliable bids. Owners need this index to get the likely cost of the project, while contractors use it to submit their financial offers in the tendering phase. The expense of building is complex. Materials rates, labour capital, and other expenses tend to fluctuate. This economic instability may have a major effect on the business, particularly on long-term and mega projects. Clear consideration is needed to limit potential financial risks and then delay the project. Construction cost indices are used to predict expected costs [142].

3.8 Using of RapidMiner

The using of RapidMiner software for prediction the EAC consists of the following steps until achieve the prediction data as following:

1- Open the programme as a first step then click on New process for creating the required modelling as shown in Figure3.15.

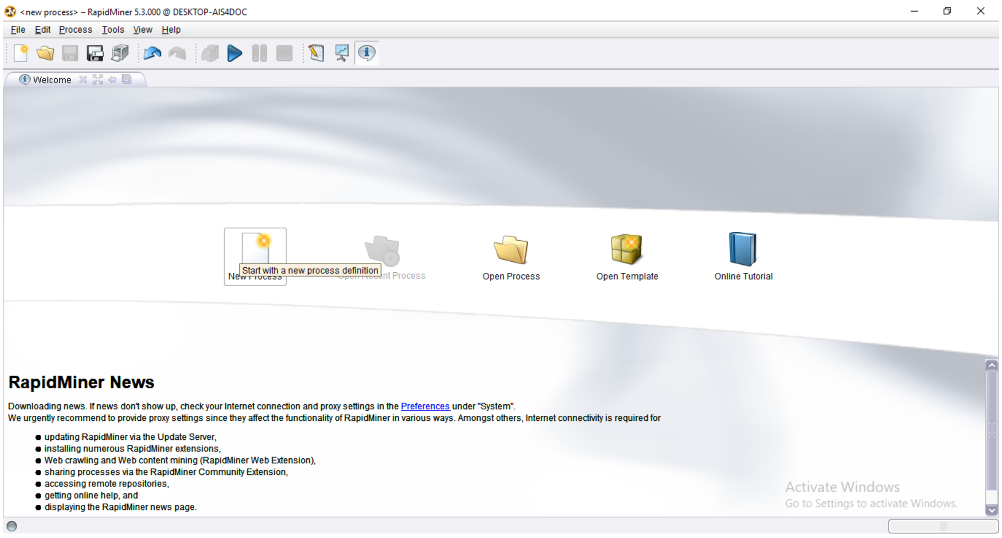
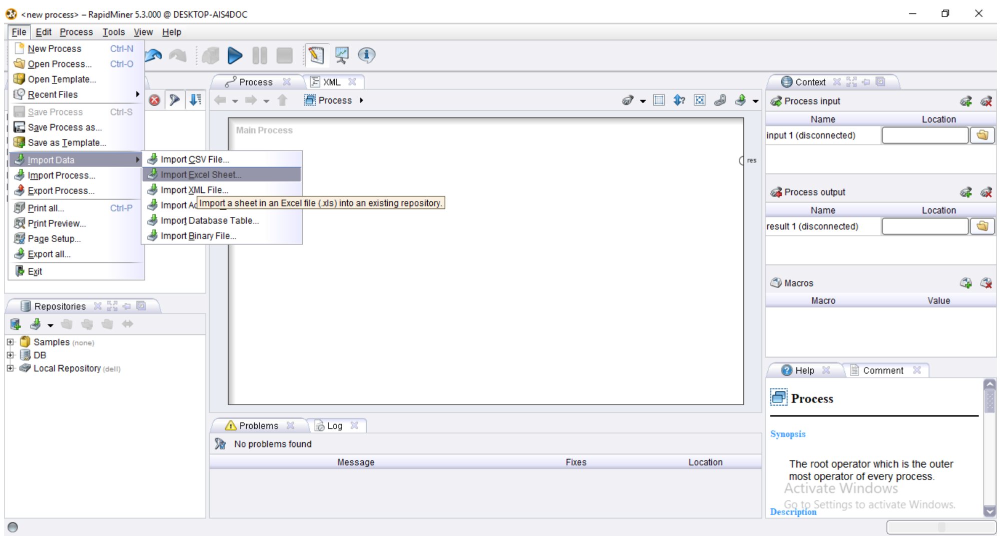


Figure 3.15 RapidMiner main page.

2- Starting the insert of projects data as well demonstrate in Figure 3.16.



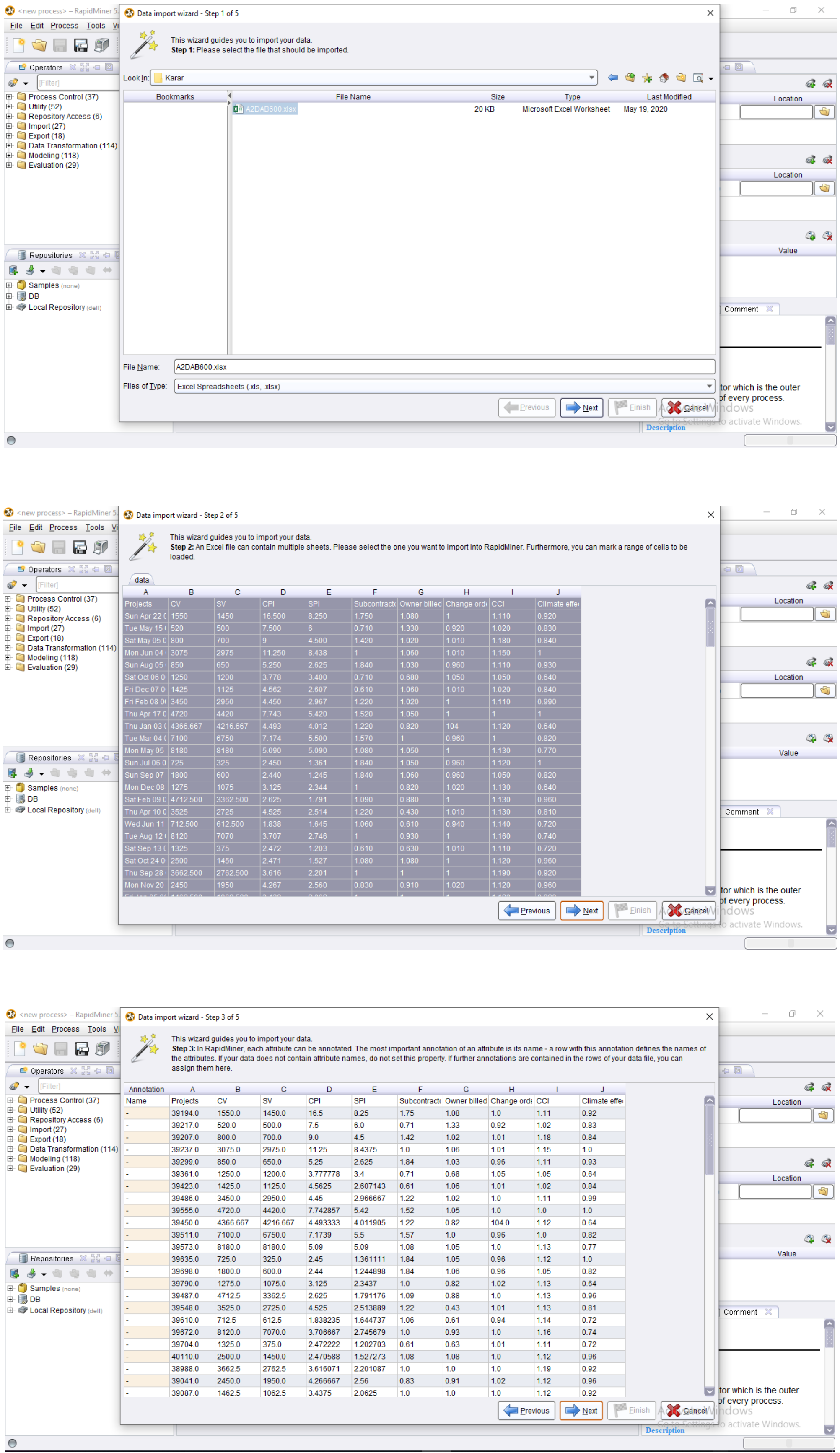


Figure 3. 16 The steps inserting of data.

3- Switch to design perspective and click on the operation and select the required samples to complete the whole mode as shown in Figure 3.17 and 3.18.

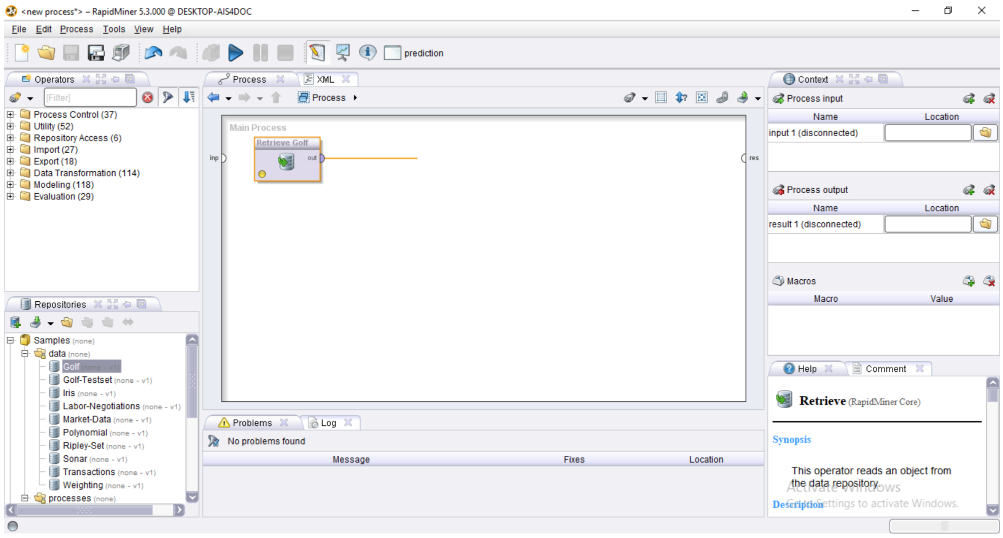


Figure 3.17 Start Building the model.

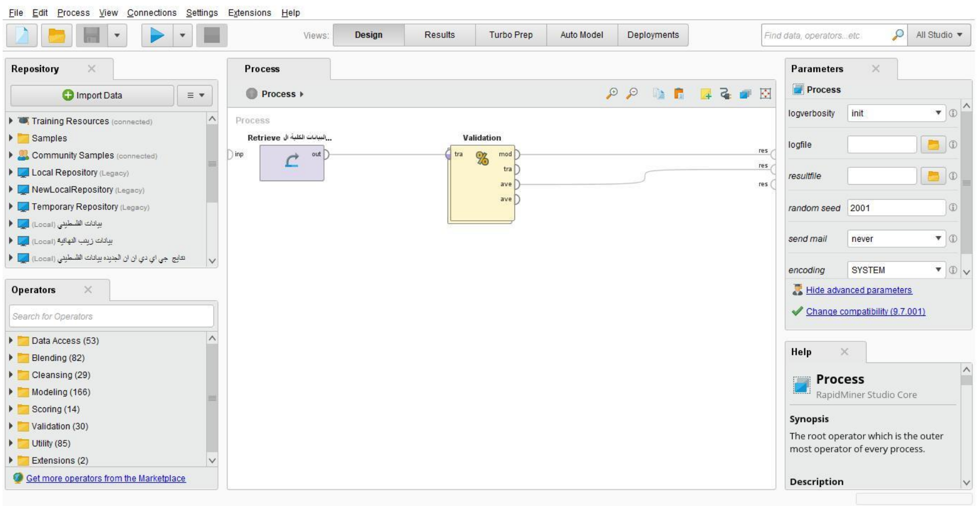


Figure 3.18 The whole SVR Prediction Model.

4- By clicking the process button and Run, the process of prediction will start, and the result of prediction will be presented on the results screen.

The previous steps just for SVR or other algorthim prediction, but the same steps could be used immediately after building the required model of production.

The modelled historical data are processed through a linear normalization scale between (0 and 1). This is for the purpose to supply the data for the programming environment with scaled numerical. The normalization is performed as followed [143]:

$$x_{new} = \left(\frac{x - x_{min}}{x_{max} - x_{min}}\right) \tag{3.7}$$

where $x_{new}$is the normalized value of the calculated EAC, x is the observed EAC, x_{max} and x_{min} are the maximum and minimum values of the EAC. The predictive models are examined using several numerical indicators that present the absolute error evaluation (the closest to zero) the best-goodness (the closest to one). In that way, more justification can be done on the optimal model for the best input combination. The numerical indicators are root mean square error (RMSE) [144], mean absolute error (MAE), mean relative error (MRE), Nash-Sutcliffe coefficient (NSE) [145], scatter index (SI), Willmott's index (WI) [146] [147]. Both the root mean square error (RMSE) and the mean absolute error (MAE) are regularly employed in model evaluation studies. Willmott and Matsuura [148] have suggested that the RMSE is not a good indicator of average model performance and might be a misleading indicator of average error. Thus the MAE would be a better metric for that purpose. While some concerns over using RMSE raised by Willmott and Matsuura [148] and Willmott et al. [149] are valid, the proposed avoidance of RMSE in favour of MAE is not the solution. Citing the papers as mentioned earlier, many researchers chose MAE over RMSE to present their model evaluation statistics when presenting or adding the RMSE measures could be more beneficial. The RMSE is more appropriate to represent model performance than the MAE when the error distribution is expected to be Gaussian. Besides, we show that the RMSE satisfies the triangle inequality requirement for a distance metric, where Willmott et al. [149] indicated that the sums-of-squares-based statistics do not satisfy this rule. In the end, we discussed some circumstances where using the RMSE will be more beneficial. However, we do not contend that the RMSEis superior over the MAE. Instead, a combination of metrics, including but certainly not limited to RMSEs and MAEs, are often required to assess model performance. The mathematical can be described as followed:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(EAC_a - EAC_p)^2}{n}} \tag{3.8}$$

$$MAE = \frac{\sum_{i=1}^{n}\left|EAC_a - EAC_p\right|}{n} \tag{3.9}$$

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{EAC_a - EAC_p}{EAC_a}\right) \tag{3.10}$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}(EAC_a - EAC_p)^2}{\sum_{i=1}^{n}(EAC_a - \overline{EAC_a})^2}\right] \tag{3.11}$$

$$SI = \frac{\sqrt{\frac{\sum_{i=1}^{n}(EAC_a - EAC_p)^2}{n}}}{\overline{EAC_a}} \tag{3.12}$$

$$WI = 1 - \left[\frac{\sum_{i=1}^{n}(EAC_a - EAC_p)^2}{\sum_{i=1}^{n}(\left|EAC_p - \overline{EAC_a}\right| + |EAC_a - \overline{EAC_a}|)^2}\right] \tag{3.13}$$

$$RE = \left[\frac{EAC_a - EAC_p}{EAC_a}\right] * 100 \tag{3.14}$$

where EAC_a is the actual observation, EAC_p is the predicted value, $\overline{EAC_a}$ and $\overline{EAC_p}$ are the mean values of the actual and predicted value.

4

RESULTS AND DISCUSSION

4.1 Introduction

As an advanced stage for the prediction process, the cost database of the selected projects was determined. The data represents the planned and the actual cost values for each month and the computed difference between them. The mathematical relationship between the nine (the abstracted input combinations) attributes and the EAC is explored using the potential of the AI expertise learning. The motivation of applying the AI models in computing the EAC is to overcome the drawbacks of the classical indexed formulations since AI models can mimic the human brain intelligence in solving complex real-life problems.

4.2 Stand-alone Predictive Models and Results

The primary prediction modelling conducted for the stand-alone proposed DNN and its comparable SVR predictive model. Table 4.1 tabulated the performance prediction skills indicators using all nine declared variables. It is observable that DNN outperformed the SVR model through prediction skills. In quantitative terms, DNN attained (i.e., RMSE-MAE) and (i.e., NSE-WI) are (0.0365-0.0264) and (0.5592-0.7531); whereas, SVR attained the prediction indicators as (0.0460-0.0333) and (0.2999-0.7114). There is a notable augmentation between the proposed and predominated data-intelligence SVR predictive model. Scatter plot was reported to explain the variation between the actual and predicted EAC for the stand-alone DNN and SVR models. A scatter plot is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data by using the excel sheet, as shown in a Figure 4.1. The points are coded (colour/shape/size); one additional variable can be displayed. The data are displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis [150]. In this study used the scatter plot used to present Predicted EAC in the y-axis and Actual data in the x-axis. Where the redpoint represent the relation between Predicted EAC and Actual data for (43) case training.

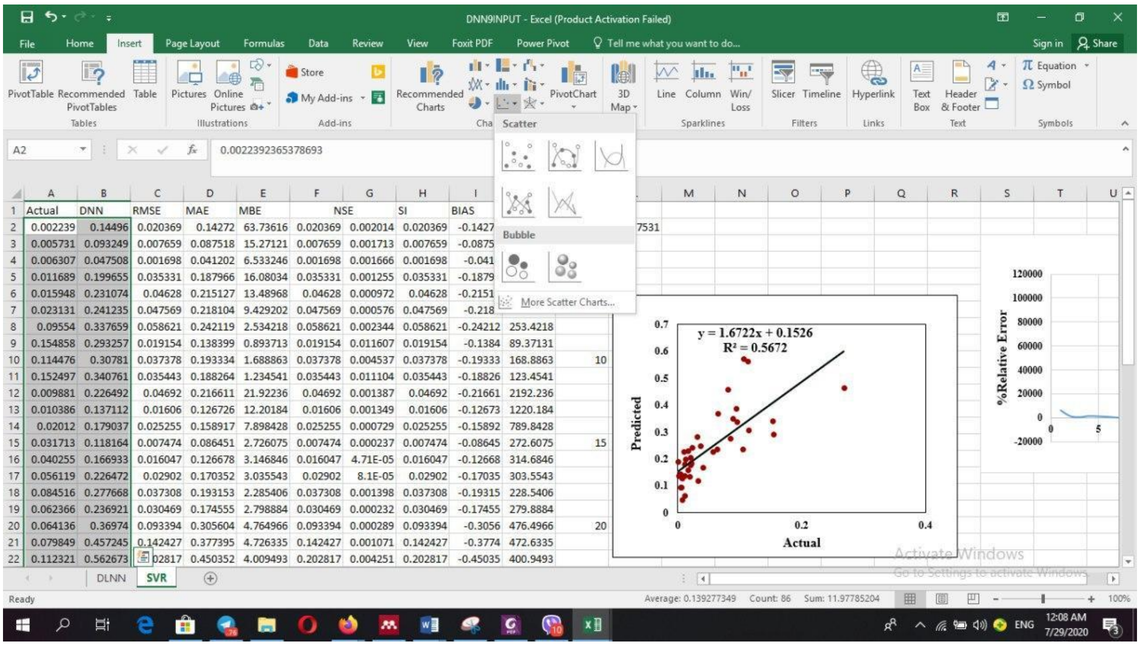


Figure 4.1 A screenshot for preparing to scatter plot by excel sheet.

Table 4.1 The numerical evaluation indicators for the DNN and SVR predictive models "stand-alone versions" over the testing modelling phase

Predictive models	RMSE	MAE	MRE	NSE	SI	WI
DNN	0.0365	0.0264	4.7680	0.5592	0.7760	0.7531
SVR	0.0460	0.0333	-7.660	0.2999	0.9780	0.7114

Figure (4.2) demonstrates (scatter plot) for the relation between predicted EAC and actual by selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The first graph refers to using DNN model to represent the Variables, R^2 value was 0.5672 and line formula (y = 0.6271x + 0.0198), while Figure (4.3) refer to using SVR model to represent the Variables, R^2 = 0.5061 and line formula (y = 0.7225x-0.0055), 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. From figures (4.2

and 4.3), it could be observed that the using of DNN model give a better determination coefficient (R^2) than the SVR model.

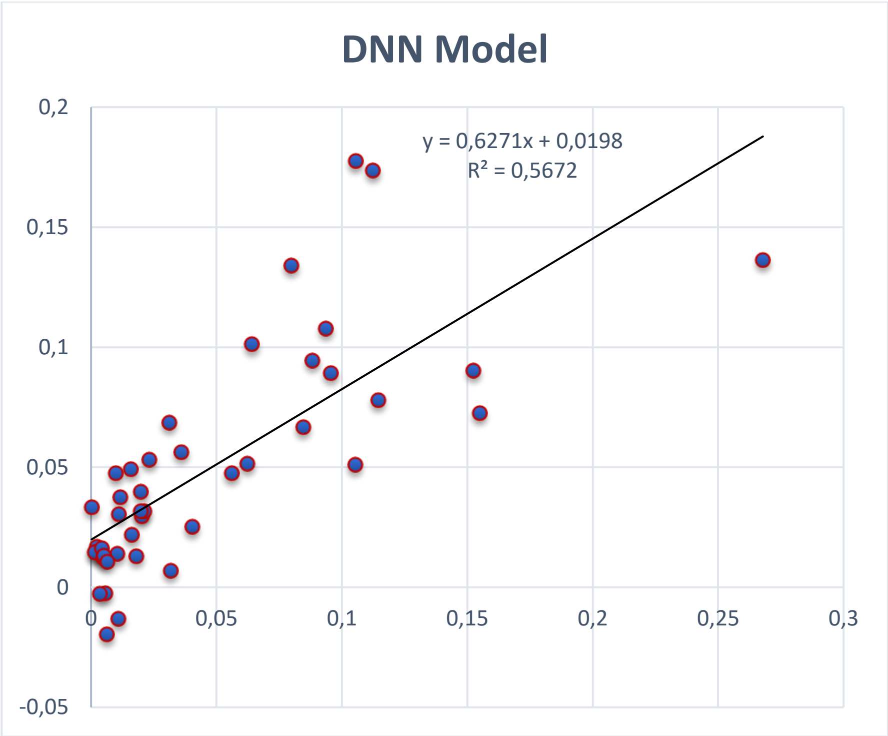


Figure 4.2 The scatter graphical plot visualization between the actual observation of EAC for the stand-alone intelligence predictive DNN models.

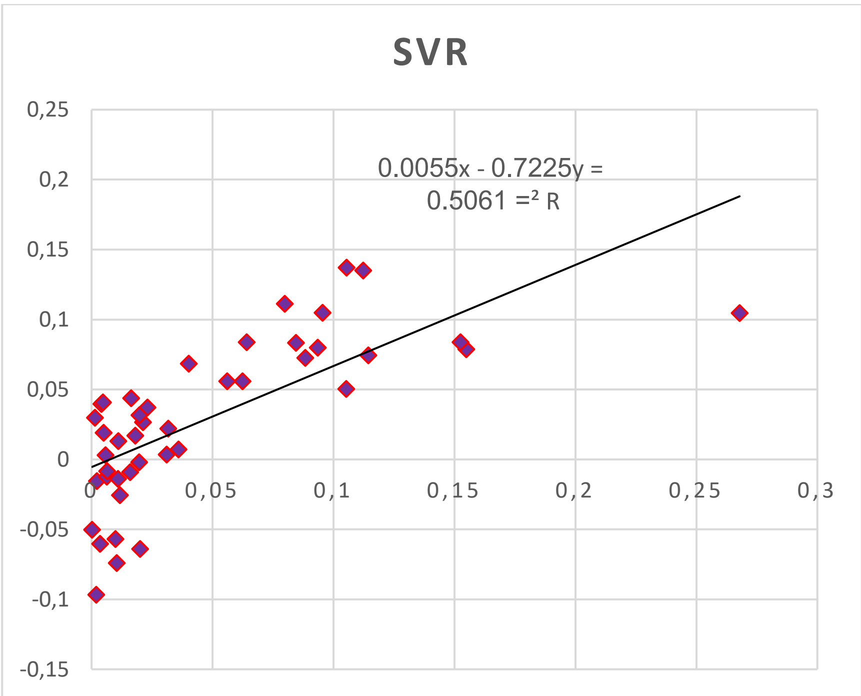


Figure 4.3 The scatter graphical plot visualization between the actual observation of EAC for the stand-alone intelligence predictive SVR models.

4.3 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 nature-inspired genetic algorithm hybridized with the DNN to abstract the suitable input combination. On the other hand, the brute-force selection procedure is used as a benchmark for the GA comparison.

The input combination and the prediction skill results of the hybrid model GA-DNN are indicated in Tables 4.2 and 4.3, respectively. By studying the archived modelling results in Table 4.3, Model 2 exhibited the excellent input combination for predicting EAC through including (CV, SV and CCI) variables as inputs for the prediction matrix. The results showed minimum absolute error metrics (e.g. RMSE-MAE) (0.0281-0.0199) and best-fit-goodness (e.g., NSE-WI) (0.7394-0.8602). The hybrid BF-DNN model behaved differently (Tables 4.4 and 4.5), seven input variables represented in CV, SV, CPI, SPI, Subcontractor billed index, Owner billed index, and Change order index, gave the optimal prediction skills with minimum RMSE ≈ 0.0280 and WI ≈ 0.8653. Note that BF-DNN surpassed the capability of the GA-DNN model; however, with more features for comprehending the internal mapping relationship between predictors and predicted.

Scatter plot graphical exhibition is one of the excellent ways to visualize the correlation between the actual observations and predicted value. Figures 4.3-4.6 presented the diversion from the ideal line of the 45^{o}. The presentation showed a noticeable agreement for the hybrid DNN over the hybrid SVR model.

Figure (4.4) demonstrates the correlation between predicted actual EAC and GA-DNN predictive models for model 1 with (CV and CCI) **Variables.** Besides, by selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using of GA-DNN give R^2 value = 0.7219 and line formula (y = 0.7782x + 0.011).

Table 4.2 The input combination attributes used to determine the value of the EAC using GA-DNN model

No.of inputs	Models	Input Variables	
2 inputs	**Model 1**	CV, CCI	EAC
3 inputs	**Model 2**	CV, SPI, CCI	EAC
4 inputs	**Model 3**	CV, CPI, Subcontractor billed index, CCI	EAC
5 inputs	**Model 4**	CV, SPI, Subcontractor billed index, Owner billed index, Climate effect index	EAC
6 inputs	**Model 5**	CV, SPI, Subcontractor billed index, Owner billed index, CCI, Climate effect index	EAC
7 inputs	**Model 6**	CV, SV, SPI, Subcontractor billed index, Owner billed index, CCI, Climate effect index	EAC
8 inputs	**Model 7**	CV, SV, CPI, SPI, Subcontractor billed index, Owner billed index, Change order index, Climate effect index	EAC

Table 4.3 The numerical evaluation indicators for the GA-DNN predictive model over the testing modelling phase

Method	RMSE	MAE	MRE	NSE	SI	WI
Model 1	0.0293	0.0213	2.3340	0.7174	0.6214	0.8496
Model 2	**0.0281**	**0.0199**	**1.5474**	**0.7394**	**0.5968**	**0.8602**
Model 3	0.0288	0.0209	3.4690	0.7262	0.6117	0.8560
Model 4	0.0294	0.0201	3.8066	0.7149	0.6241	0.8473
Model 5	0.0285	0.0201	-0.5278	0.7322	0.6049	0.8570
Model 6	0.0292	0.0209	5.5415	0.7191	0.6196	0.8483
Model 7	0.0380	0.0232	4.4742	0.5244	0.8061	0.7355

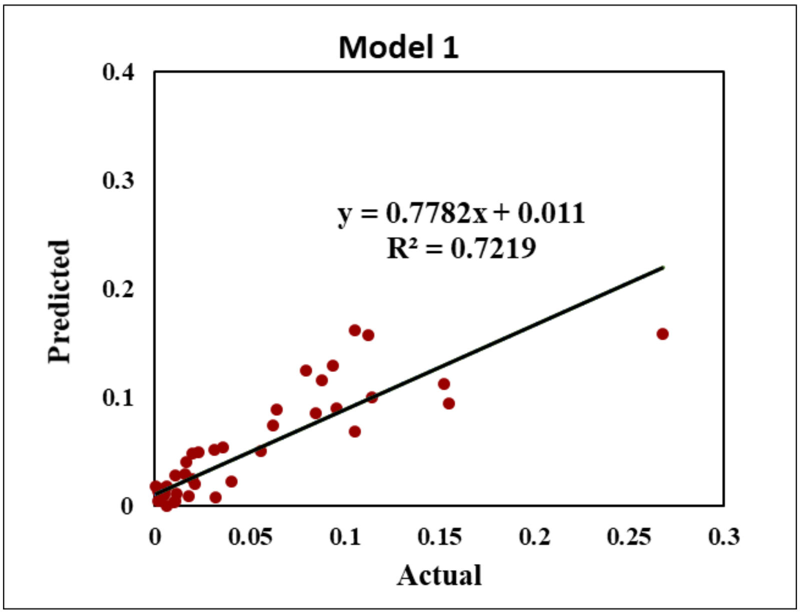


Figure 4.4 The (scatter plot) graphical visualization for the intelligence predictive GA-DNN model 1.

Figure (4.5) demonstrates the correlation between predicted actual EAC and GA-DNN predictive models for model 2 with (CV, SPI, and CCI) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using of GA-DNN give R^2 value = 0.7399 and line formula (y = 0.7581x + 0.0119).

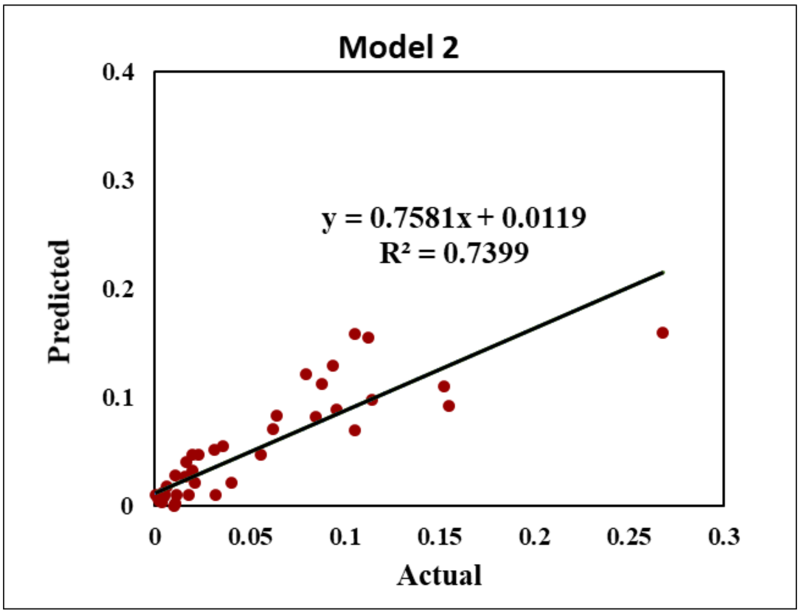


Figure 4.5 The (scatter plot) graphical visualization for the intelligence predictive GA-DNN model 2.

Figure (4.6) demonstrates the correlation between predicted actual EAC and GA-DNN predictive models for model 3 with (CV, CPI, Subcontractor billed index,and CCI) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using of GA-DNN give R^2 value = 0.7327 and line formula (y = 0.757x + 0.0119).

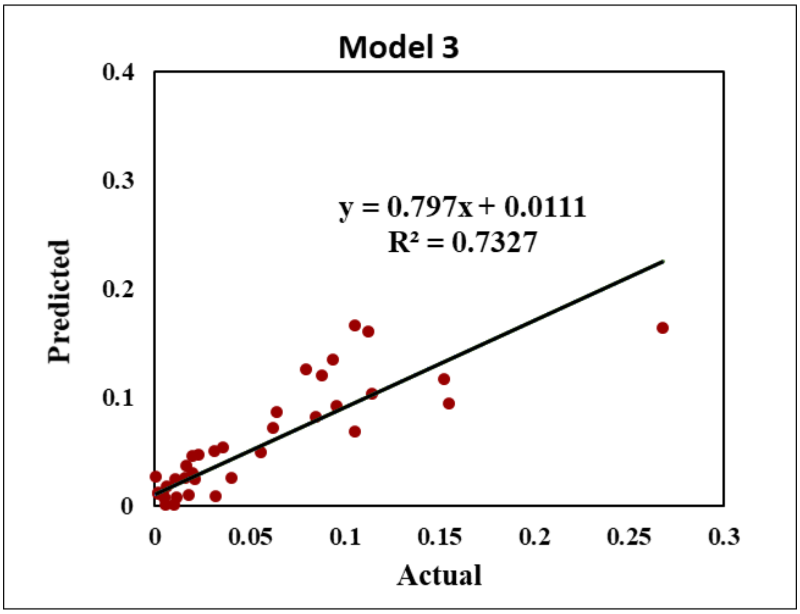


Figure 4.6 The (scatter plot) graphical visualization for the intelligence predictive GA-DNN model 3.

Figure (4.7) demonstrates the correlation between predicted actual EAC and GA-DNN predictive models for model 4 with (CV, SPI, Subcontractore billed index, Owner billed index and Climate effect index) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using of GA-DNN give R^2 value = 0.718 and line formula (y = 0.7495x + 0.014).

Figure (4.8) demonstrates the correlation between predicted actual EAC and GA-DNN predictive models for model 5 with (CV,SPI,Subcontractore billed index,Owner billed index, CCI and Climate effect index) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using of GA-DNN give R^2 value = 0.7345 and line formula (y = 0.7755x + 0.0108).

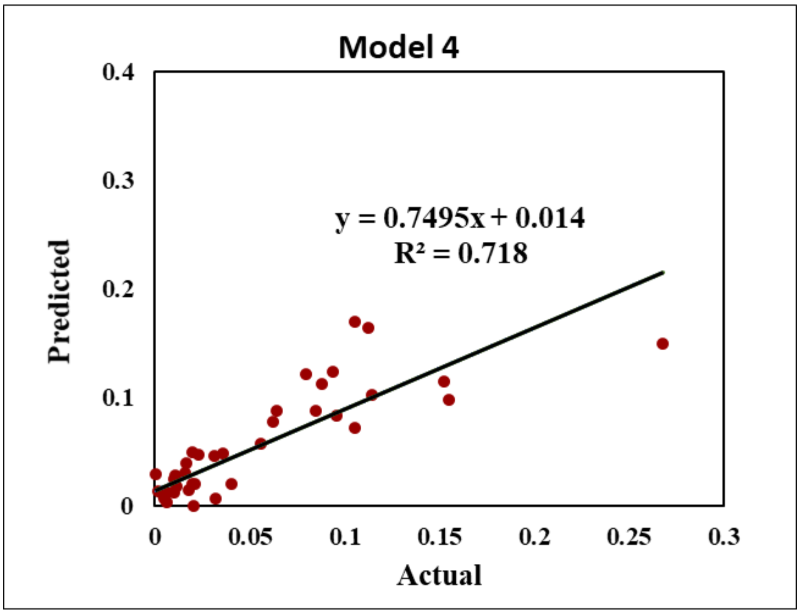


Figure 4.7 The (scatter plot) graphical visualization for the intelligence predictive GA-DNN model 4.

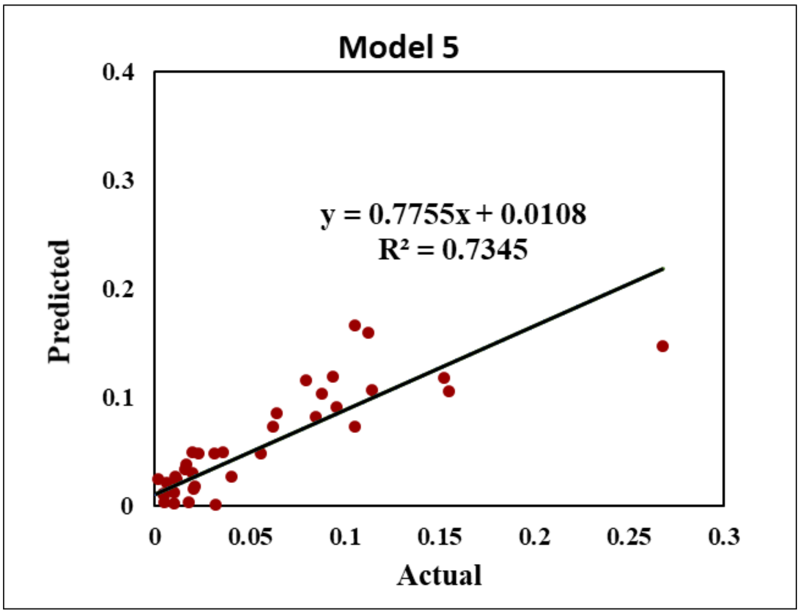


Figure 4.8 The (scatter plot) graphical visualization for the intelligence predictive GA-DNN model 5.

Figure (4.9) demonstrates the correlation between predicted actual EAC and GA-DNN predictive models for model 6 with (CV, SV, SPI, Subcontractor billed index, Owner billed index, CCI,and Climate effect index) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using of GA-DNN give R^2 value = 0.7196 and line formula (y = 0.7374x + 0.0118).

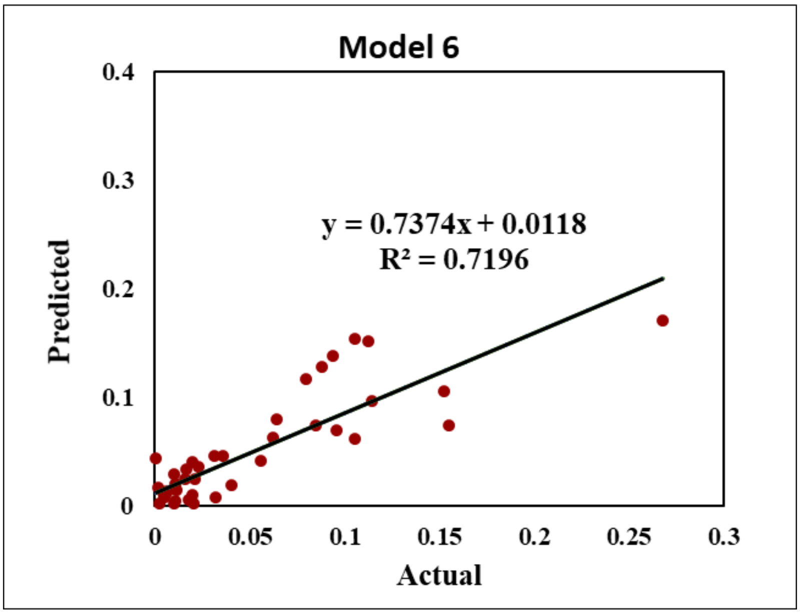


Figure 4.9 The (scatter plot) graphical visualization for the intelligence predictive GA-DNN model 6.

Figure (4.10) demonstrates the correlation between predicted actual EAC and GA-DNN predictive models for model 7 with (CV, SV, CPI, SPI, Subcontractor billed index, Owner billed index, Change order index and Climate effect index) **Variables.**

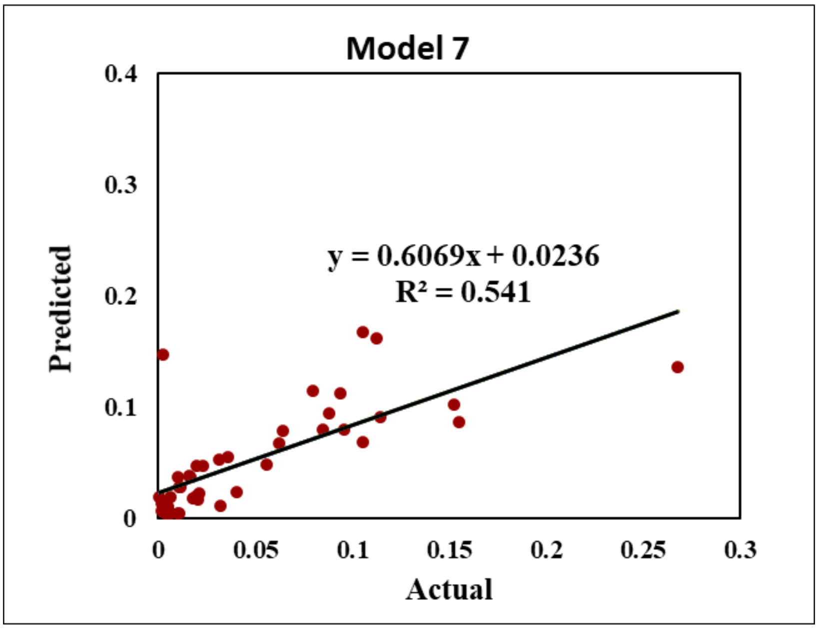


Figure 4.10 The (scatter plot) graphical visualization for the intelligence predictive GA-DNN model 7.

By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using of GA-DNN give R^2 value = 0.541 and line formula (y = 0.6069x + 0.0236)

By using GA-DNN model, model seven with eight variables gives the lowest R^2 value = 0.541, while model 2 with just three variables give the highest WI value = 0.8602.

Table 4.4 The input combination attributes used to determine the value of the EAC using BF-DNN model

No. of inputs	Models	Input Variables	
2 inputs	**Model 1**	CV, SV	EAC
3 inputs	**Model 2**	CV, SV, CPI	EAC
4 inputs	**Model 3**	CV, SV, CPI, SPI	EAC
5 inputs	**Model 4**	CV, SV, CPI, SPI, Subcontractor billed index	EAC
6 inputs	**Model 5**	CV, SV, CPI, SPI, Subcontractor billed index, Owner billed index	EAC
7 inputs	**Model 6**	CV, SV, CPI, SPI, Subcontractor billed index, Owner billed index, Change order index	EAC
8 inputs	**Model 7**	CV, SV, CPI, SPI, Subcontractor billed index, , Owner billed index, Change order index, CCI	EAC

Table 4.5 The numerical evaluation indicators for the BF-DNN predictive model over the testing modelling phase

Models	RMSE	MAE	MRE	NSE	SI	WI
Model 1	0.0308	0.0202	2.1798	0.6880	0.6530	0.8335
Model 2	0.0333	0.0202	0.6291	0.6335	0.7076	0.8024
Model 3	0.0281	0.0185	1.7501	0.7395	0.5966	0.8634
Model 4	0.0297	0.0181	-0.0110	0.7101	0.6294	0.8472
Model 5	0.0306	0.0189	2.0955	0.6923	0.6484	0.8369
Model 6	**0.0280**	**0.0171**	**0.7838**	**0.7425**	**0.5932**	**0.8653**
Model 7	0.0314	0.0188	1.2414	0.6747	0.6667	0.8281

The modelling input combinations and prediction skills results of the GA-SVR and BF-SVR tabulated in Table 4.6-4.9. In comparison with GA-SVR model, GA-DNN demonstrated a remarkable enhancement in term of the quantitative units measurable (RMSE-MAE) are decreased by (26.3-20.1%); whereas, (NSE-WI) are augmented by (8.8-4.2%). This proved the capability of the GA-DNN model on mimicking the actual relationship of the project elements on the EAC phenomena.

Figure (4.11) demonstrates the correlation between predicted actual EAC and BF-DNN predictive models for model 1 with (CV and SV) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using of BF-DNN give R^2 value = 0.6948 and line formula (y = 0.7599x + 0.0099).

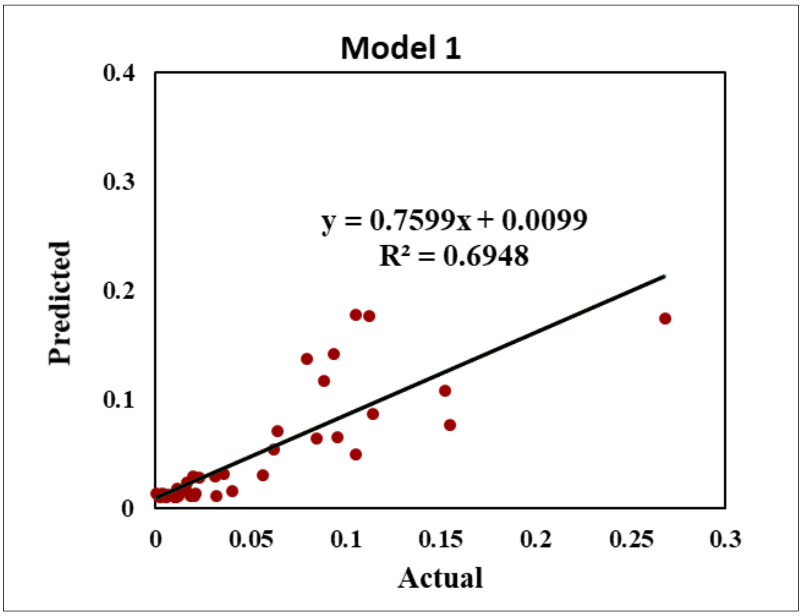


Figure 4.11 The (scatter plot) graphical visualization for the intelligence predictive BF-DNN model 1.

Figure (4.12) demonstrates the correlation between predicted actual EAC and BF-DNN predictive models for model 2 with (CV, SV, and CPI) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using of BF-DNN give R^2 value = 0.6439 and line formula (y = 0.649x + 0.011).

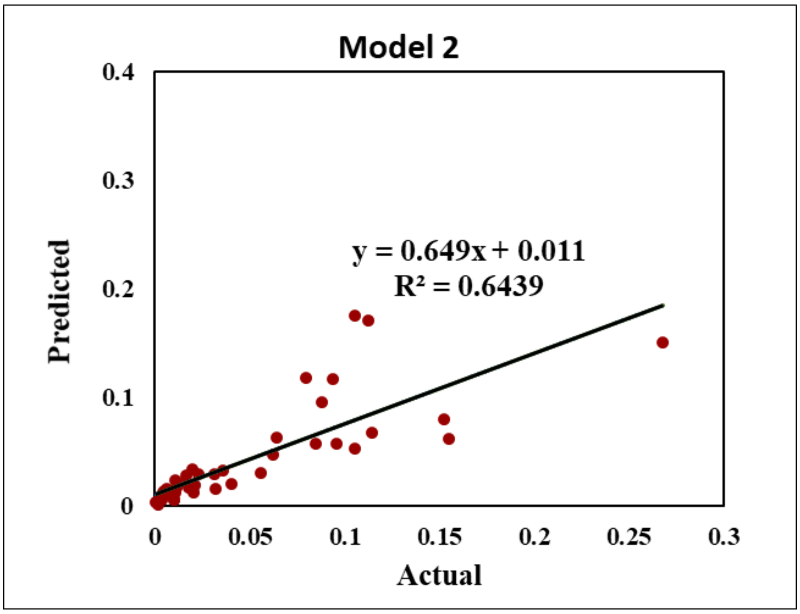


Figure 4.12 The (scatter plot) graphical visualization for the intelligence predictive BF-DNN model 2.

Figure (4.13) demonstrates the correlation between predicted actual EAC and BF-DNN predictive models for model 3 with (CV, SV, CPI, and SPI) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using of BF-DNN give R^2 value = 0.7455 and line formula (y = 0.8155x + 0.0097).

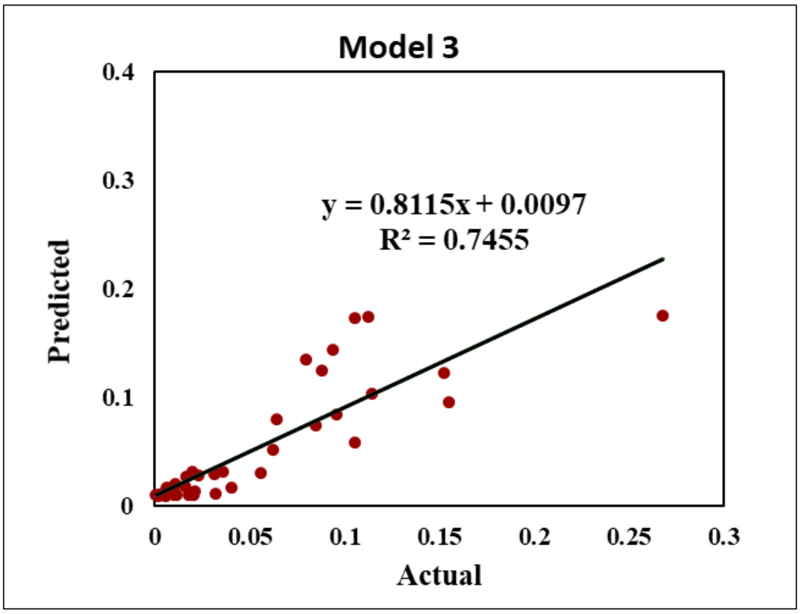


Figure 4.13 The (scatter plot) graphical visualization for the intelligence predictive BF-DNN model 3.

Figure (4.14) demonstrates the correlation between predicted actual EAC and BF-DNN predictive models for model 4 with (CV, SV, CPI, SPI, and Subcontractor billed index) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using of BF-DNN give R^2 value = 0.7177 and line formula (y = 0.6998x + 0.0095).

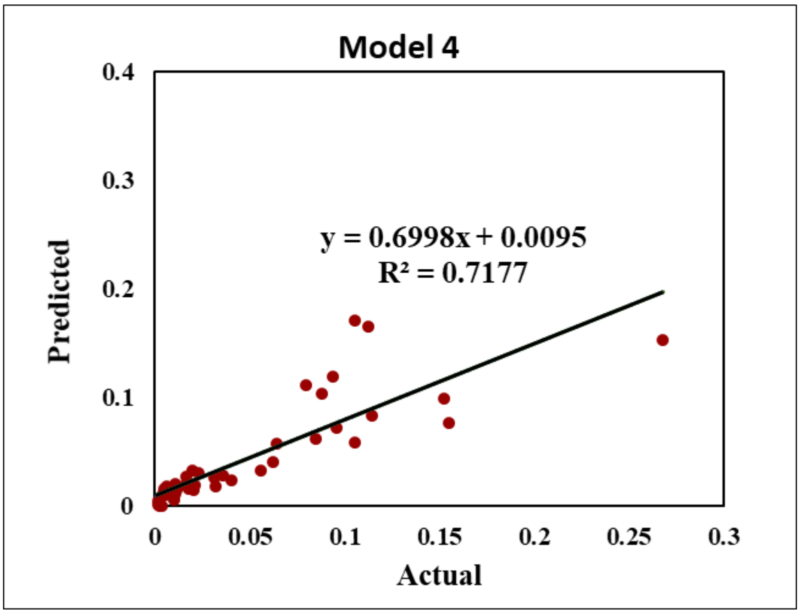


Figure 4.14 The (scatter plot) graphical visualization for the intelligence predictive BF-DNN model 4.

Figure (4.15) demonstrates the correlation between predicted actual EAC and BF-DNN predictive models for model 5 with (CV, SV, CPI, SPI, Subcontractor billed index, and Owner billed index) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using of BF-DNN give R^2 value = 0.7004 and line formula (y = 0.7424x + 0.008).

Figure (4.16) demonstrates the correlation between predicted actual EAC and BF-DNN predictive models for model 6 with (CV, SV, CPI, SPI, Subcontractor billed index, Owner billed index, and Change order index) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using of BF-DNN give R^2 value = 0.7487 and line formula (y = 0.8066x + 0.0068).

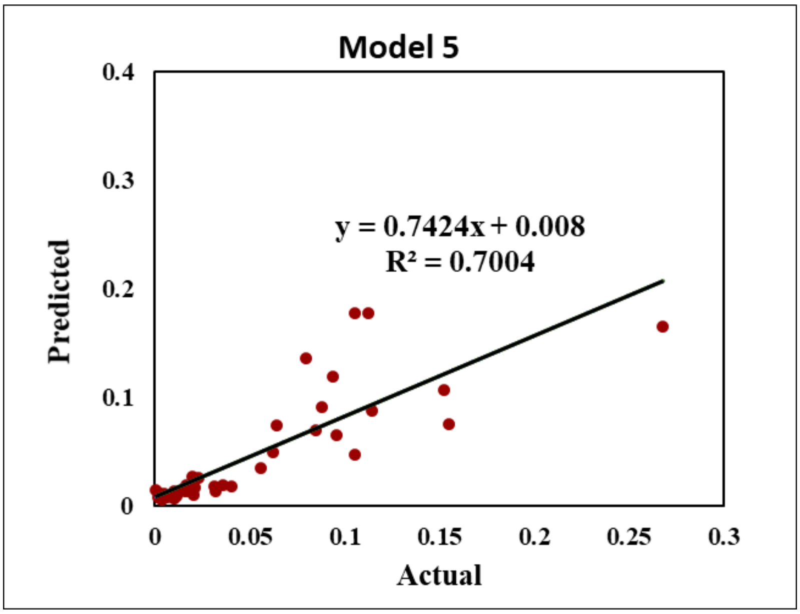


Figure 4. 15 The (scatter plot) graphical visualization for the intelligence predictive BF-DNN model 5.

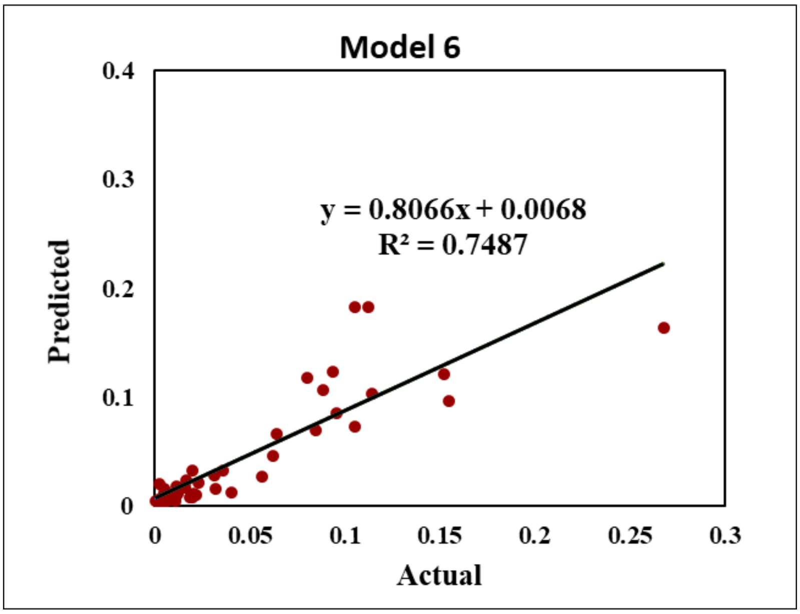


Figure 4.16 The (scatter plot) graphical visualization for the intelligence predictive BF-DNN model 6.

Figure (4.17) demonstrates the correlation between predicted actual EAC and BF-DNN predictive models for model 7 with (CV, SV, CPI, SPI, Subcontractor billed index, Owner billed index, Change order index, Climate effect index) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using of BF-DNN give R^2 value = 0.6858 and line formula (y = 0.6812x + 0.0092).

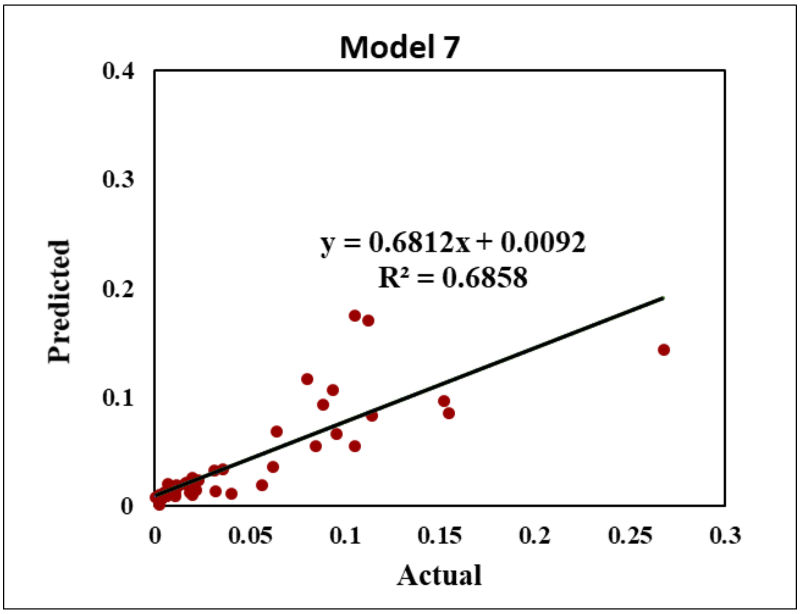


Figure 4.17 The (scatter plot) graphical visualization for the intelligence predictive BF-DNN model 7.

By using BF -DNN model, model 2 with three variables gives the lowest R^2 value = 0.6439, while model 6 with seven variables give the highest R^2 value = 0.7487. Therefore, using of GA-DNN gives better results compared with BF -DNN models.

Table 4.6 The input combination attributes used to determine the value of the EAC using GA-SVR model

No. of inputs	Models	The Type of Input Variables	
2 inputs	**Model 1**	CV, CCI	EAC
3 inputs	**Model 2**	CV, Subcontractor billed index, CCI	EAC
4 inputs	**Model 3**	CV, SV, SPI, Subcontractor billed index	EAC
5 inputs	**Model 4**	CV, CPI, SPI, Subcontractor billed index, CCI	EAC
6 inputs	**Model 5**	CV, CPI, SPI, Subcontractor billed index, CCI, Climate effect index	EAC
7 inputs	**Model 6**	CV, CPI, SPI, Subcontractor billed index, Owner billed index, CCI, Climate effect index	EAC
8 inputs	**Model 7**	CV, CPI, SPI, Subcontractor billed index, Owner billed index, Change order index, CCI, Climate effect index	EAC

Table 4. 7 The numerical evaluation indicators for the GA-SVR predictive model over the testing modeling phase

Method	RMSE	MAE	MRE	NSE	SI	WI
Model 1	0.0827	0.0548	-0.0126	0.6827	0.3671	0.8350
Model 2	0.0851	0.0552	0.0034	0.6642	0.3776	0.8204
Model 3	0.0921	0.0556	-0.1181	0.6065	0.4088	0.8271
Model 4	0.0816	0.0586	0.0992	0.6911	0.3622	0.8323
Model 5	**0.0807**	**0.0560**	**0.0726**	**0.6983**	**0.3579**	**0.8370**
Model 6	0.0818	0.0591	0.1216	0.6900	0.3628	0.8335
Model 7	0.0847	0.0538	0.0501	0.6674	0.3759	0.8301

Figure (4.18) demonstrates the correlation between predicted actual EAC and GA-SVR predictive models for model 1 with (CV and CCI) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using GA-SVR give R^2 value = 0.6973 and line formula (y = 0.6456x + 0.0646).

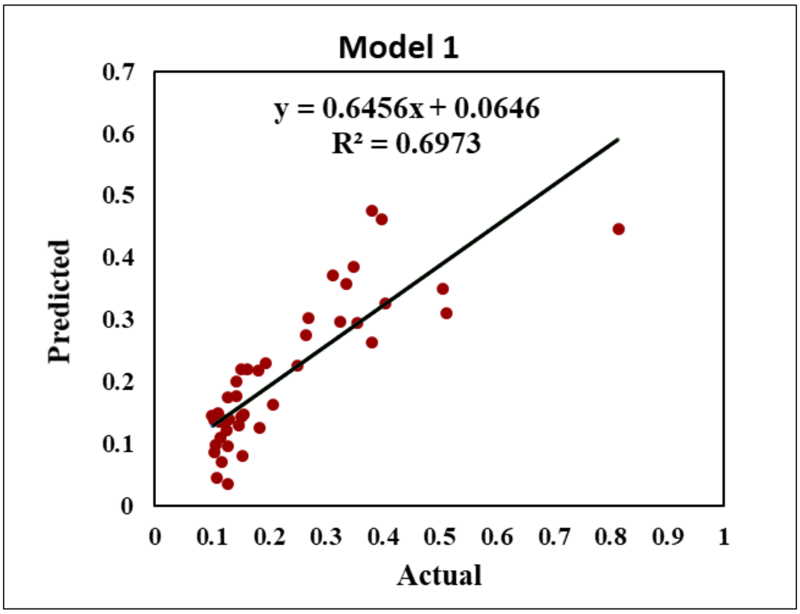


Figure 4.18 The (scatter plot) graphical visualization for the intelligence predictive GA-SVR model 1.

Figure (4.19) demonstrates the correlation between predicted actual EAC and GA-SVR predictive models for model 2 with (CV, Subcontractor billed index, and CCI)

Variables. By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using GA-SVR give R^2 value = 0.6731 and line formula (y = 0.6391x + 0.0689).

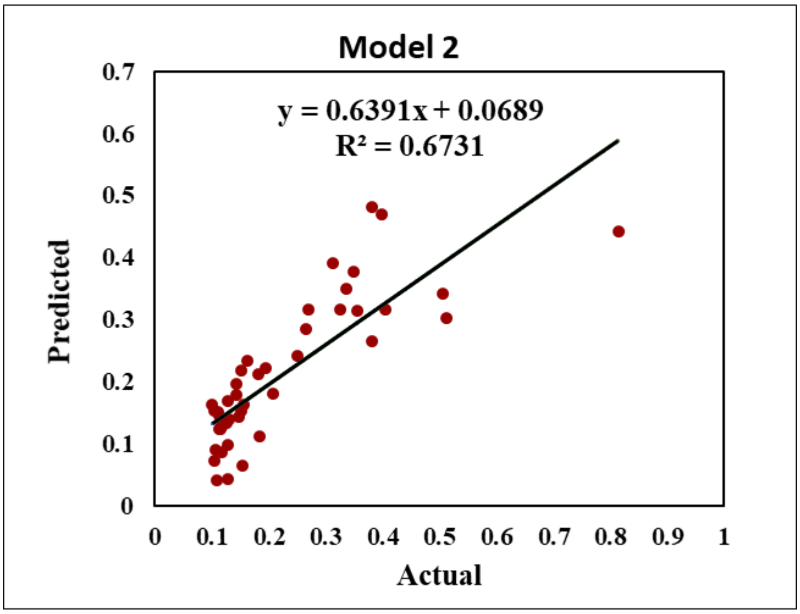


Figure 4.19 The (scatter plot) graphical visualization for the intelligence predictive GA-SVR model 2.

Figure (4.20) demonstrates the correlation between predicted actual EAC and GA-SVR predictive models for model 3 with (CV, SV, SPI, and Subcontractore billed index) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using GA-SVR give R^2 value = 0.6841 and line formula (y = 0.5815x + 0.0577).

Figure (4.21) demonstrates the correlation between predicted actual EAC and GA-SVR predictive models for model 4 with (CV, CPI, SPI, Subcontractor billed index and CCI) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using GA-SVR give R^2 value = 0.6927 and line formula (y = 0.6958x + 0.0744).

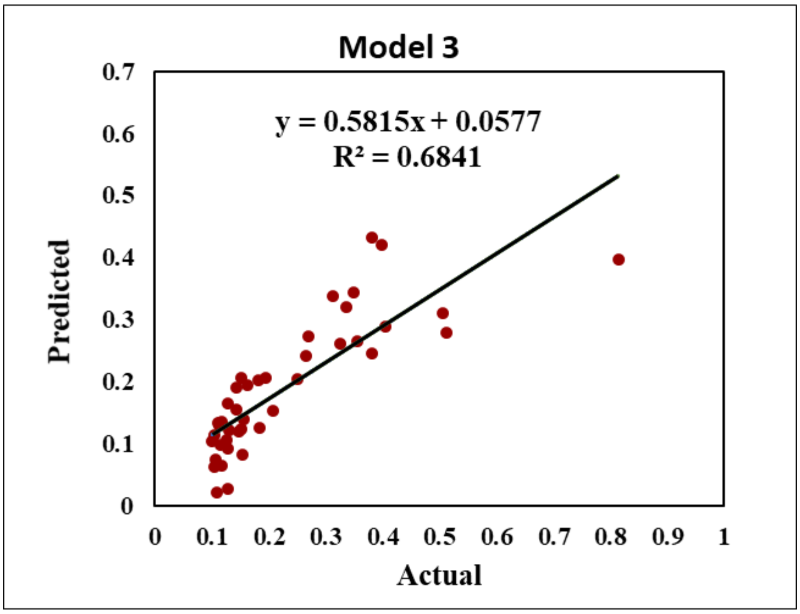


Figure 4.20 The (scatter plot) graphical visualization for the intelligence predictive GA-SVR model 3.

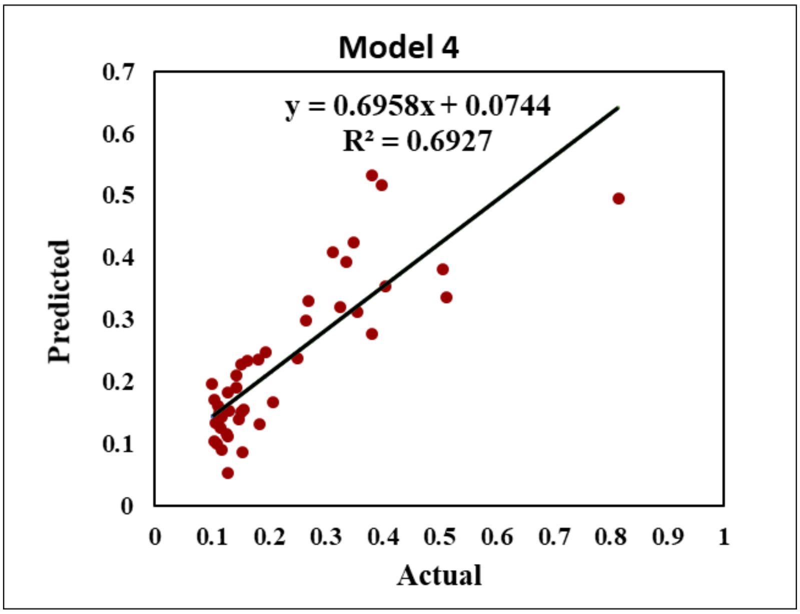


Figure 4.21 The (scatter plot) graphical visualization for the intelligence predictive GA-SVR model 4.

Figure (4.22) demonstrates the correlation between predicted actual EAC and GA-SVR predictive models for model 5 with (CV, CPI, SPI, Subcontractor billed index, CCI, and Climate effect index) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using GA-SVR give R^2 value = 0.7005 and line formula (y = 0.6616x + 0.0764).

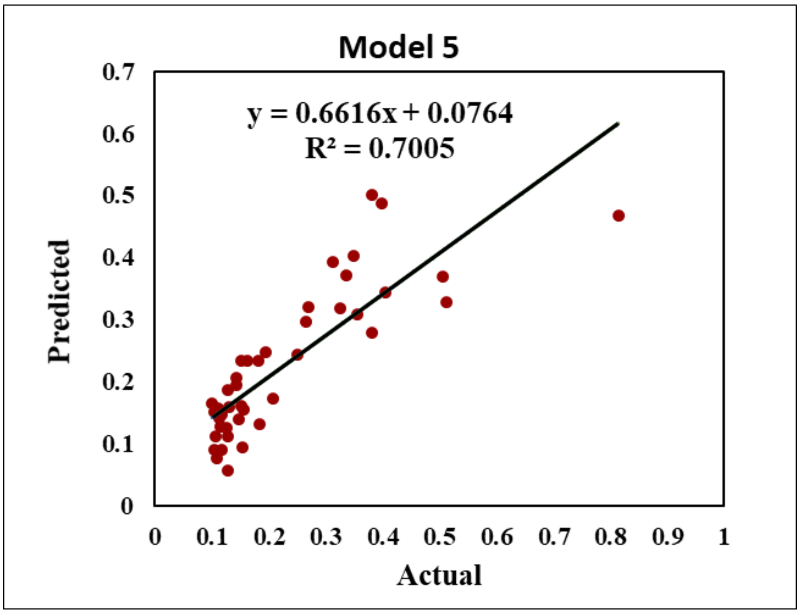


Figure 4.22 The (scatter plot) graphical visualization for the intelligence predictive GA-SVR model 5.

Figure (4.23) demonstrates the correlation between predicted actual EAC and GA-SVR predictive models for model 6 with (CV, CPI, SPI, Subcontractore billed index, Owner billed index ,CCI, and Climate effect index) **Variables.**

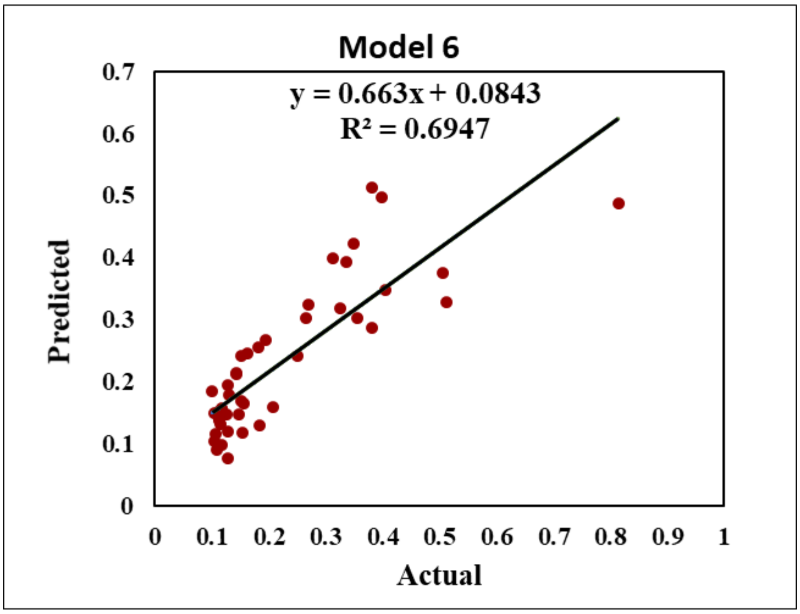


Figure 4.23 The (scatter plot) graphical visualization for the intelligence predictive GA-SVR model 6.

By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using GA-SVR give R^2 value = 0.6947 and line formula (y = 0.663x + 0.0843).

Figure (4.24) demonstrates the correlation between predicted actual EAC and GA-SVR predictive models for model 7 with (CV, CPI, SPI, Subcontractor billed index, Owner billed index, Change order index , CCI, and Climate effect index) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using GA-SVR give R^2 value = 0.689 and line formula (y = 0.5797x + 0.0851).

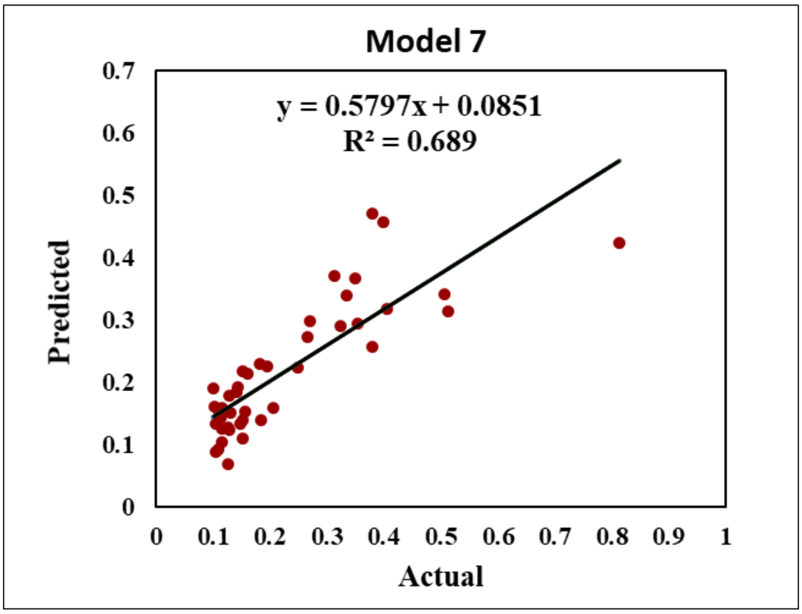


Figure 4.24 The (scatter plot) graphical visualization for the intelligence predictive GA-SVR model 7.

By using GA -SVR model, model two with three variables gives the lowest WI value = 0.8204, while model 5 with just six variables give the highest WI value =0.8370. Therefore, using of GA-DNN give better results comparison with GA -SVR models.

Table 4.8 The input combination attributes used to determine the value of the EAC using BF-SVR model

No. of inputs	Models	The Type of Input Variables	
2 inputs	**Model 1**	CV, SV	EAC
3 inputs	**Model 2**	CV, SV, CPI	EAC
4 inputs	**Model 3**	CV, SV, CPI, SPI	EAC
5 inputs	**Model 4**	CV, SV, CPI, SPI, Subcontractor billed index	EAC
6 inputs	**Model 5**	CV, SV, CPI, SPI, Subcontractor billed index, Owner billed index	EAC
7 inputs	**Model 6**	CV, SV, CPI, SPI, Subcontractor billed index, Owner billed index, Change order index	EAC
8 inputs	**Model 7**	CV, SV, CPI, SPI, Subcontractor billed index, Owner billed index, Change order index, CCI	EAC

Table 4.9 The numerical evaluation indicators for the BF-SVR predictive model over the testing modeling phase

Method	RMSE	MAE	MRE	NSE	SI	WI
Model 1	0.0975	0.0648	0.0501	0.6477	0.3868	0.8247
Model 2	0.0958	0.0719	-0.0368	0.5772	0.4238	0.8240
Model 3	0.0971	0.0794	0.3066	0.5627	0.4309	0.8165
Model 4	0.1023	0.0820	0.3898	0.5149	0.4539	0.8258
Model 5	0.0972	0.0740	0.1753	0.6058	0.4092	0.8117
Model 6	**0.0957**	**0.0735**	**0.1110**	**0.5755**	**0.4246**	**0.8304**
Model 7	0.0963	0.0600	0.1660	0.6133	0.4052	0.7923

Figure (4.25) demonstrates the correlation between predicted actual EAC and BF-SVR predictive models for model 1 with (CV and SV) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using BF-SVR give R^2 value = 0.6801 and line formula (y = 0.8227x + 0.0473).

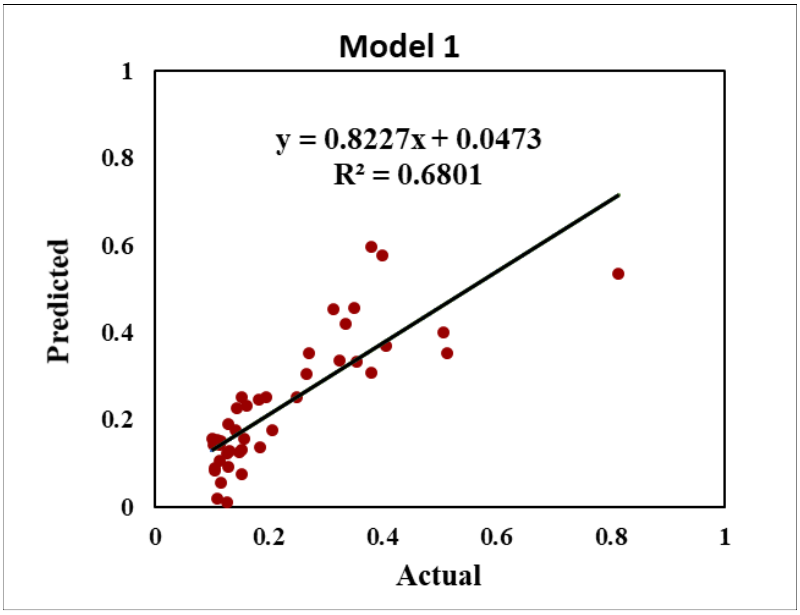


Figure 4.25 The (scatter plot) graphical visualization for the intelligence predictive BF-SVR model 1.

Figure (4.26) demonstrates the correlation between predicted actual EAC and BF-SVR predictive models for model 2 with (CV, SV, and CPI) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using BF-SVR give R^2 value = 0.6789 and line formula (y = 0.9416x + 0.0129).

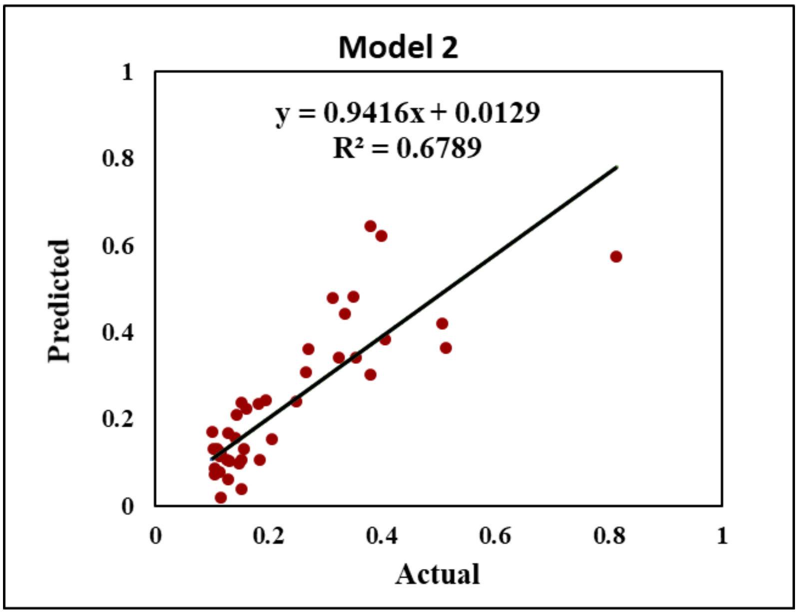


Figure 4.26 The (scatter plot) graphical visualization for the intelligence predictive BF-SVR model 2.

Figure (4.27) demonstrates the correlation between predicted actual EAC and BF-SVR predictive models for model 3 with (CV, SV, CPI, and SPI) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using BF-SVR give R^2 value = 0.6666 and line formula (y = 0.7217x + 0.109).

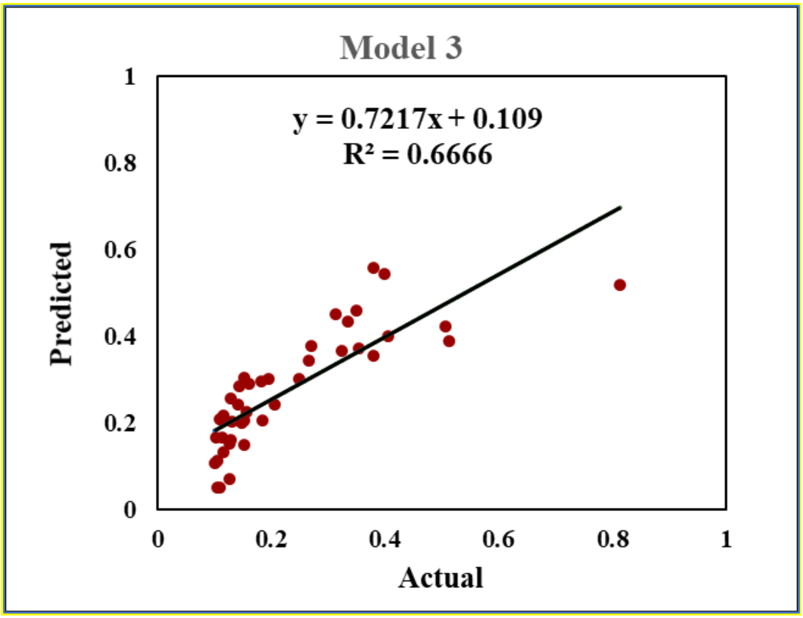


Figure 4.27 The (scatter plot) graphical visualization for the intelligence predictive BF-SVR model 3.

Figure (4.28) demonstrates the correlation between predicted actual EAC and BF-SVR predictive models for model 4 with (CV, SV, CPI, SPI, and Subcontractor billed index) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using BF-SVR give R^2 value = 0.682 and line formula (y = 0.7328x + 0.1196).

Figure (4.29) demonstrates the correlation between predicted actual EAC and BF-SVR predictive models for model 5 with (CV, SV, CPI, SPI, Subcontractor billed index, and Owner billed index) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using BF-SVR give R^2 value = 0.6588 and line formula (y = 0.7715x + 0.0785).

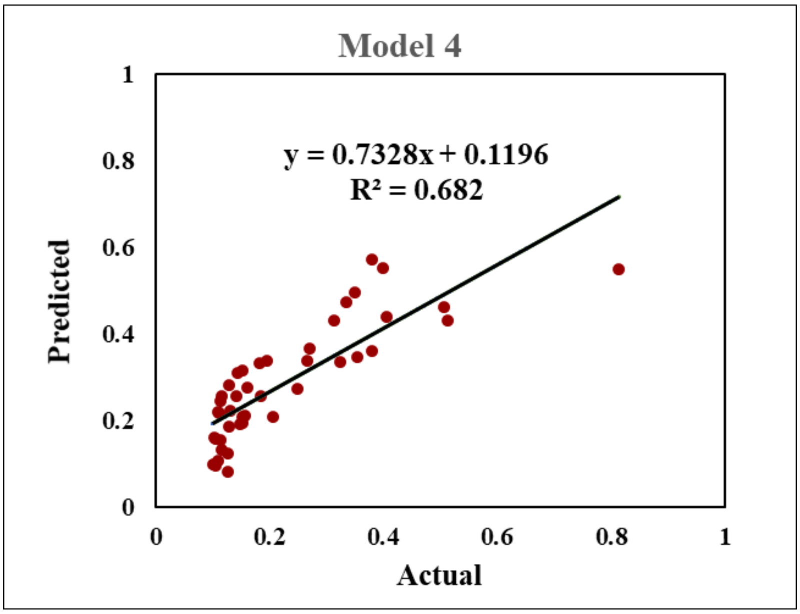


Figure 4.28 The (scatter plot) graphical visualization for the intelligence predictive BF-SVR model 4.

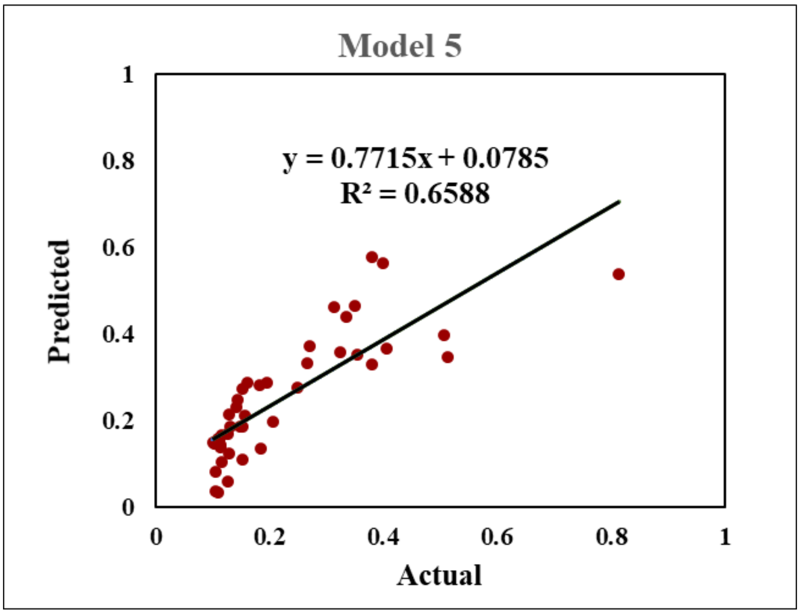


Figure 4.29 The (scatter plot) graphical visualization for the intelligence predictive BF-SVR model 5.

Figure (4.30) demonstrates the correlation between predicted actual EAC and BF-SVR predictive models for model 6 with (CV, SV, CPI, SPI, Subcontractor billed index, Owner billed index, and Change order index) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using BF-SVR give R^2 value = 0.6895 and line formula (y = 0.9345x + 0.0389).

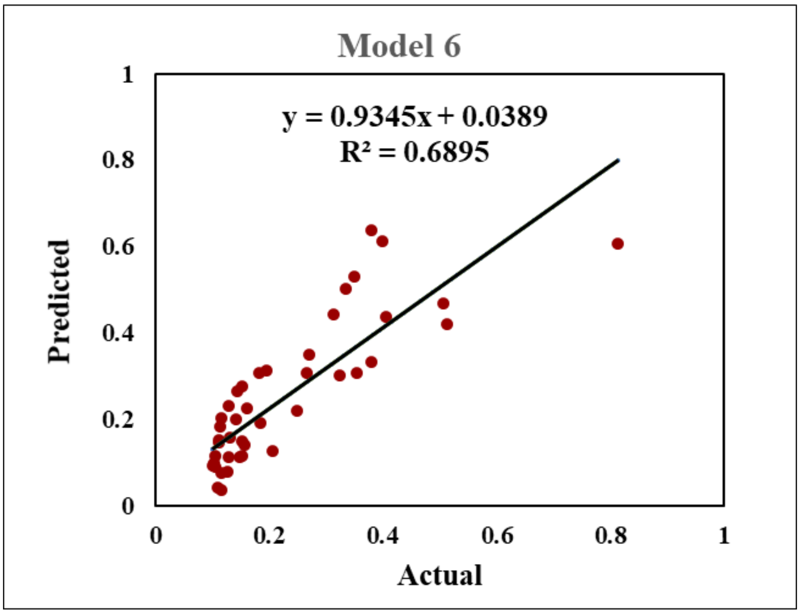


Figure 4.30 The (scatter plot) graphical visualization for the intelligence predictive BF-SVR model 6.

Figure (4.31) demonstrates the correlation between predicted actual EAC and BF-SVR predictive models for model 7 with (CV, SV, CPI, SPI, Subcontractor billed index, Owner billed index, Change order index, and CCI) **Variables.** By the selective trials utilizing equations (3.8, 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14) to achieve the best model expression for the coefficient of determination (R^2) value and the line formula. The using BF-SVR give R^2 value = 0.6278 and line formula (y = 0.5405x + 0.1106).

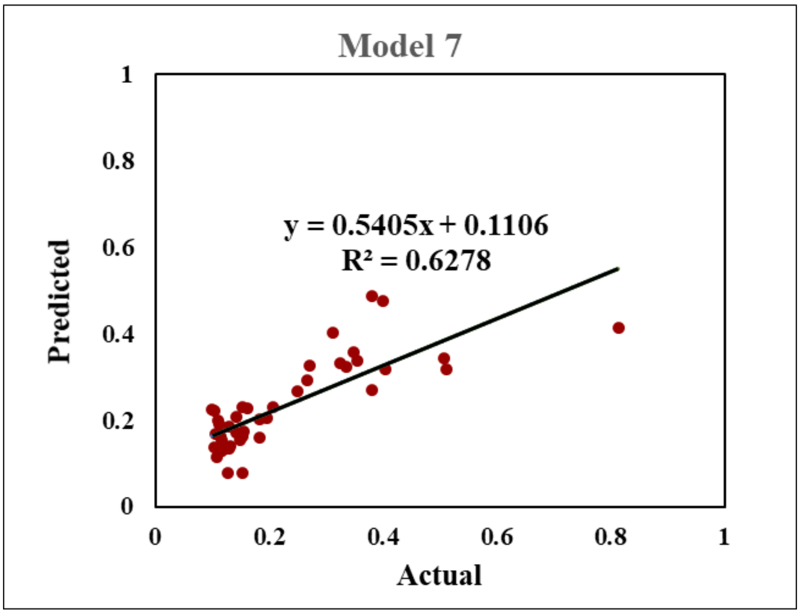


Figure 4.31 The (scatter plot) graphical visualization for the intelligence predictive BF-SVR model 7.

By using BF-SVR model, model six with seven variables gives the highest R^2 value = 0.6895, while model 5 with six variables give the highest R^2 value = 0.7005. Therefore, using of GA-SVR give better results comparison with BF - SVR models.

Finally, Figure 4.32 and 4.33 revealed the testing phase of the modelling for all the established predictive models. GA-DNN and BF-DNN showed a noticeable matching with the actual EAC. Figure (4.32) show the relationship between testing phase EAC prediction and observation and prediction EAC values for observed EAC, GA-DNN and GA-SVR, in most cases, GA-DNN gives closest results to observed EAC better than GA-SVR. Figure (4.33) show the relationship between testing phase EAC prediction and observation and prediction EAC values for observed EAC, BF-DNN and BF-SVR, in most cases, BF-DNN gives closest results to observed EAC better than BF-SVR.

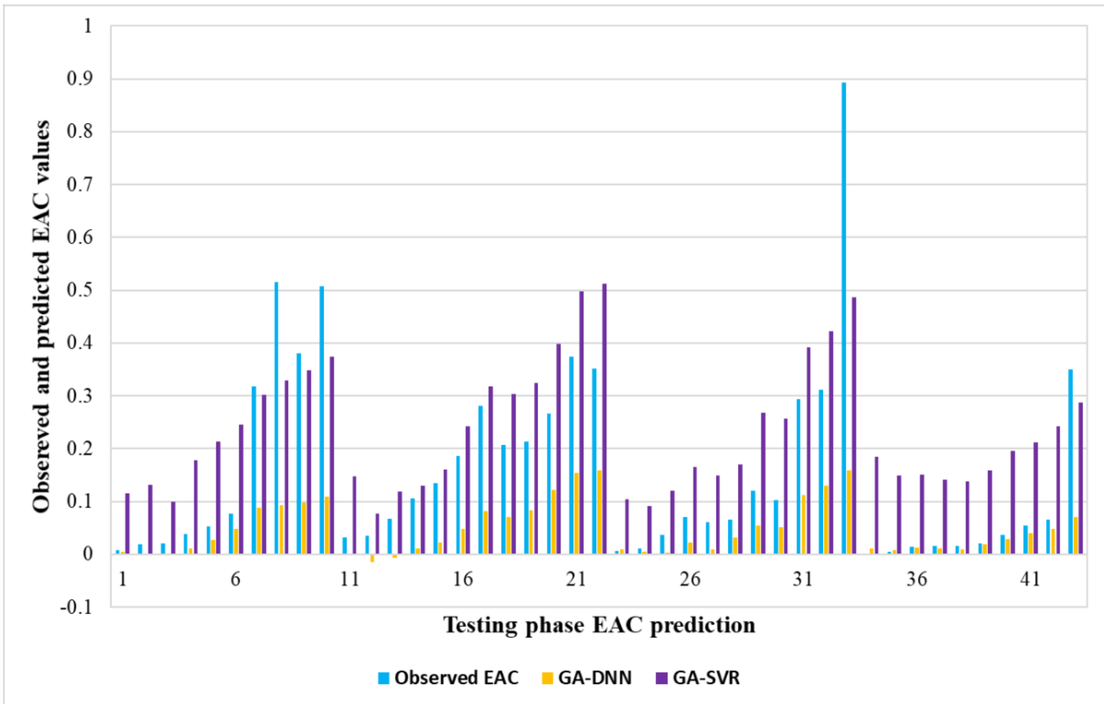


Figure 4.32 Actual observation of EAC and the optimal combination for GA-DNN and GA-SVR

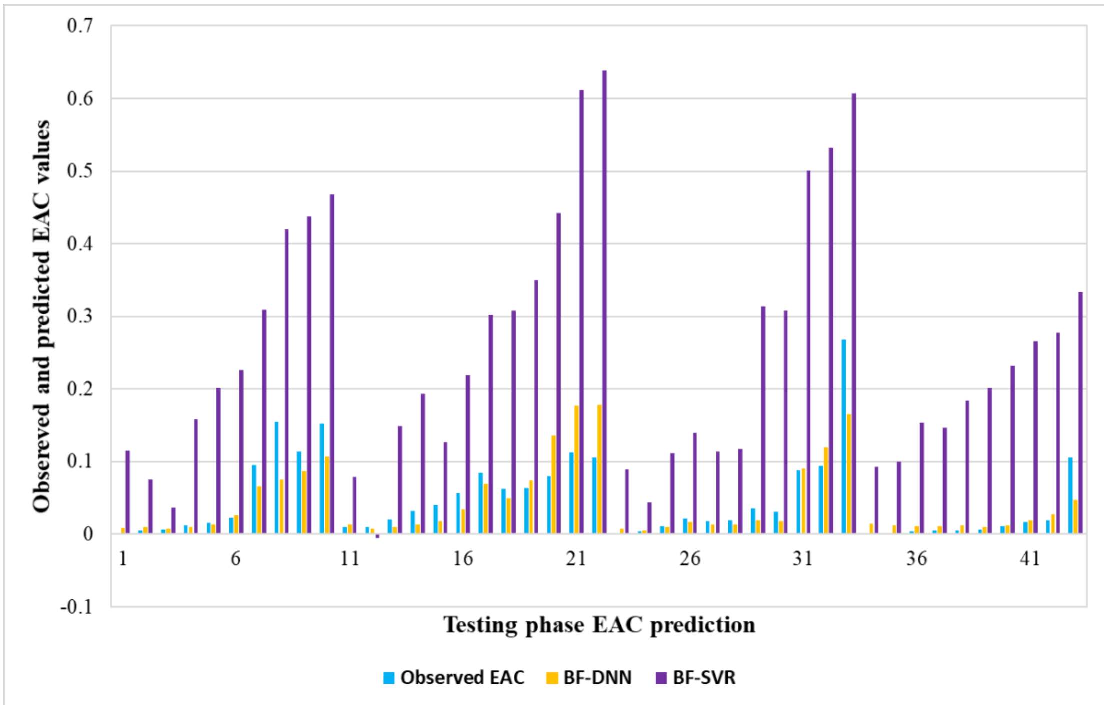


Figure 4.33 Actual observation of EAC and the optimal combination for BF-DNN and BF-SVR

4.4 Discussion Summary

The applied methodology in the current research was inspired by the motivation of exploring a new reliable approach for modelling EAC in construction projects. The proposed model distinguished itself by the capability of comprehending the actual mechanism of the related variables to the targeted variable with more solidity manners. This is the main essential perspective for practical implementation from construction project management. Overall, the hybridization of the evolutionary optimization algorithm as a selective procedure pre-predictive model (i.e., deep neural network) attained convincing results for the perspective of the scientific research and innovative modelling strategy exploration.

The number of input data does not affect the fit-of-goodness, for example, GA -SVR model, model seven with eight variables (CV, CPI, SPI, Subcontractor billed index, Owner billed index, Change order index, CCI, and Climate effect index) does not gives the highest R2 value = 0.689 it is approximately the lowest one with approximately the highest RMSE = 0.0847. As well as using BF -DNN model, model seven with eight variables (CV, SV, CPI, SPI, Subcontractor billed index, , Owner billed index, Change order index, and CCI) gives approximately the lowest R2 value = 0.6858 with approximately the highest RMSE =0.0314. The same indications were observed by using GA-DNN model, model seven with eight variables (CV, SV, CPI, SPI, Subcontractor billed index, Owner billed index, Change order index, and Climate effect index) gives the lowest R2 value = 0.541 with highest RMSE =0.0380. Based on the various statistical indicators, the best results indicated an outstanding evaluation performance concerning the minimum absolute error measures and the best fit-of-goodness (RMSE and correlation value (WI)) equal to (0.0281and 0.8602) using only three input attributes (i.e., CV, SPI, and CCI). The is evidencing the capacity of the proposed hybrid model to achieve reliable prediction accuracy with fewer input variables. It might be noticed that BF-DNN model attained performance indicators (e.g., RMSE and R2) equal to (0.0280 and 0.8653); however, this model required seven input variables information (i.e., CV, SV, CPI, SPI, Subcontractor billed index, Owner billed index and Change order index) to attain this level of accuracy. Moreover, GA-DNN model attained performance indicators (e.g., RMSE and WI) equal to (0.028 and 0.8602). From the engineering perspective, GA-DNN is more robust data-intelligence model to be implemented in actual cases as for project construction engineering, not all the time information are available and allocating such possibility of model give more credit for the engineering prospects. İn spite of the best values or both models approximately the same, but the required variables for achieving that results is different, where BF-DNN need seven viariables, while GA-DNN used just three.

5
CONCLUSION

5.1 Conclusion

Construction projects can be exceptionally complicated and are filled with uncertainty that may adversely affect their outcome. For satisfying these uncertainties, contingency and management reserve are earmarked into the budget. These and other resources are mobilized to ensure the successful completion of a project. Project success is one of the most frequently discussed topics in the field of project management, yet it is the least agreed-upon issue. Project success refers to the successful completion of cost, time, and scope objectives, quality of the project management process, and satisfying stakeholders' needs. Based on these facts, the current research is devoted. In this study, a new hybrid data-intelligence predictive model called GA-DNN is explored for facilitating the construction managers with the reliable and robust methodology that control project cost and attain accurate estimation for the estimation at completion (EAC). The implementation of this methodology is provided as an automation system where the project activities can be monitored, controlled, and any defective consequences can be avoided. The intelligence system comprises two phases: (i) the evolutionary phase of the genetic algorithm to abstract the influenced input attributes for the modelled prediction matrix, and (ii) the DNN prediction model that uses the abstracted variables for each input combination to module the EAC. BF input section procedure used as a benchmark for the GA optimizer and SVR as a comparable prediction model. For demonstrating the capability of the model in the engineering applications, historical project information obtained from fifteen projects in Iraq region is inspected in this research.

At first, the results confirmed the predictability of the DNN over the SVR stand-alone models. Besides, the hybridization with nature-inspired input algorithm selection boosted the prediction outcomes. In a quantitative presentation, the optimal prediction accuracy was reported that the minimal absolute error measures and the best fit-of-goodness (RMSE and correlation value (R^2)) equal to (0.0281 and 0.8602) using only three input parameters (i.e., CV, SPI, and CCI). This was approved the potential of the proposed hybrid model to achieve reliable prediction accuracy with fewer input variables. The results also stated that BF-DNN model attained performance indicators (e.g., RMSE and R2) equal to (0.028 and 0.8653); however, this model required seven input variables information (i.e., CV, SV, CPI, SPI, Subcontractor billed index, Owner billed index, and Change order index) to attain this level of accuracy. However, GA-DNN gave us the best result, but needs more time than the rest of the models, and this depends on the type and development of the computer, where we have applied the models in a type (processor: Intel (R) Core(TM) i3-2350M CPU @ 2.3GHz 2.30 GHz; Installed RAM: 4.00 GB; System type: 64-bit operating system, x64-based processor) computer.

The limitations of this project could be represented by the difficulties of collecting data from Iraqi construction projects due to security and political issues. In addition to the policy of construction companies to keep their construction information, mistakes and error that lead to delay the projects away from the government or other local authorities. Hence, it is not easy to give this information to other benefits. Estimation at Completion (EAC) estimates in artificial intelligent models are related to the input factors that the model was built upon, so changing in any of these inputs will increase the estimated error or make the model unusable or change the output. Therefore, it is recommended linking the model with price changes through cost index technique.

The devotion for future research is highly applicable for the current study where this methodology can be implemented on other construction projects as a real-time application where the contribution can be recognized in the form of a practical solution. It is worth to highlight, and the current investigation was experimentally developed using historical information about residential construction projects. Thus, investigating other construction projects such as industrial can be an excellent example to be examined. This can be distinguished the advantage of monitoring the project life in more reliable manners and subjective to the status of the project. Another important associated drawback is the uncertainties of data, models and input parameters that can be checked in a future study.

5.2 Recommendations for Future Work

The present project demonstrated very interesting results in predicting the Estimation at Completion (EAC) of building projects, and this method will continue to make impressive gains, particularly in civil engineering projects. Nevertheless, some recommendations must be presented for decision-makers in the construction fields and future investigation to support the results of this project;

1. All construction parties are encouraged to be more aware of Estimation at Completion (EAC) development and pay more attention to using this developed technique in the estimation process.
2. They are recommending the associations between engineering and government to establish a database for executed projects for researchers to improve Estimation at Completion (EAC) process.
3. The model should be augmented to take into consideration the other different types of construction projects, for example, infrastructure construction projects and heavy construction projects.
4. The development of artificial neural network models requires the presence of a structured database for the finished projects in the construction companies. Unfortunately, most Iraqi construction companies have no structured database system that can provide researchers with the required information. It is recommended that a standard database system for storing information regarding the finished projects should be developed and applied by the construction companies working in Iraq.
5. They are applying the same method on different projects in another country (developed countries) to investigate the differences in the construction completion between those countries and developing countries.
6. For future investigations, it is recommended to gain more training data from recent projects and inserting them to training data. This will develop the training process and produce more input choices.

REFERENCES

[1] Nassar KM, Nassar WM, Hegab MY. Evaluating cost overruns of asphalt paving project using statistical process control methods. J Constr Eng Manag 2005. https://doi.org/10.1061/(ASCE)0733-9364(2005)131:11(1173).

[2] Cheng M, Peng H, Wu Y, Chen T. Automation in Construction Estimate at Completion for construction projects using Evolutionary Support Vector Machine Inference Model. Autom Constr 2010;19:619–29. https://doi.org/10.1016/j.autcon.2010.02.008.

[3] Liu L, Zhu K. Improving cost estimates of construction projects using phased cost factors. J Constr Eng Manag 2007. https://doi.org/10.1061/(ASCE)0733-9364(2007)133:1(91).

[4] Kim GH, An SH, Kang KI. Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. Build Environ 2004. https://doi.org/10.1016/j.buildenv.2004.02.013.

[5] Navon R. Automated project performance control of construction projects. Autom Constr 2005;14:467–76. https://doi.org/10.1016/j.autcon.2004.09.006.

[6] Mahamid I. Factors contributing to poor performance in construction projects: studies of Saudi Arabia. Aust J Multi-Disciplinary Eng 2016;12:27–38. https://doi.org/10.1080/14488388.2016.1243034.

[7] Al Hallaq KAR. Causes of contractor's failure in Gaza Strip. Causes Contract Fail Gaza Strip 2003.

[8] Al-Najjar JM. Factors influencing time and cost overruns on construction projects in the Gaza Strip. Factors Influ Time Cost Overruns Constr Proj Gaza Strip 2008.

[9] Arditi D, Koksal A, Kale S. Business failures in the construction industry. Eng Constr Archit Manag 2000;7:120–32.

[10] Frimpong Y, Oluwoye J, Crawford L. Causes of delay and cost overruns in construction of groundwater projects in a developing countries; Ghana as a case study. Int J Proj Manag 2003;21:321–6.

[11] Nega F. Causes and effects of cost overrun on public building construction projects in Ethiopia 2008.

[12] Mahamid I. Factors contributing to poor performance in construction projects: studies of Saudi Arabia. Aust J Multi-Disciplinary Eng 2016;12:27–38.

[13] Long ND, Ogunlana S, Quang T, Lam KC. Large construction projects in developing countries: a case study from Vietnam. Int J Proj Manag 2004;22:553–61.

[14] Iyer KC, Jha KN. Factors affecting cost performance: evidence from Indian construction projects. Int J Proj Manag 2005;23:283–95.

[15] Faridi AS, El-Sayegh SM. Significant factors causing delay in the UAE construction industry. Constr Manag Econ 2006;24:1167–76.

[16] Kaming PF, Olomolaiye PO, Holt GD, Harris FC. Factors influencing construction time and cost overruns on high-rise projects in Indonesia. Constr Manag Econ 1997;15:83–94.

[17] Mahamid I. Contractors perspective toward factors affecting labor productivity in building construction. Eng Constr Archit Manag 2013;20:446–60. https://doi.org/10.1108/ecam-08-2011-0074.

[18] Sweis RJ, Bisharat SM, Bisharat L, Sweis G. Factors affecting contractor performance on public construction projects. Life Sci J 2014;11:28–39.

[19] Callistus T, Felix AL, Ernest K, Stephen B, Andrew AC. Factors Affecting Quality Performance of Construction Firms in Ghana: Evidence from Small–Scale Contractors. Civ Environ Res 2014;6:18–23.

[20] Fleming QW, Koppelman JM. Earned Value Project Management. Engineering 1998;16:19–23. https://doi.org/10.1016/S0263-7863(97)82251-X.

[21] Anbari FT. Earned value project management method and extensions. Proj Manag J 2003;34:12–23. https://doi.org/10.1109/EMR.2004.25113.

[22] Apolot R, Alinaitwe H, Tindiwensi D. An Investigation into the Causes of Delay and Cost Overrun in Uganda's Public Sector Construction Projects. J Constr Dev Ctries 2011. https://doi.org/10.5121/ijcsit.2011.3406.

[23] Calcutt HM. Cost growth in DoD major programs: A historical perspective. 1993.

[24] Christensen DS. Determining an accurate estimate at completion. Natl Contract Manag J 1993;25:17–25.

[25] Christensen DS. Value cost management report to evaluate the contractor's estimate. Acquis Rev Q 1999:283–96.

[26] Aliverdi R, Moslemi Naeni L, Salehipour A. Monitoring project duration and cost in a construction project by applying statistical quality control charts. Int J Proj Manag 2013. https://doi.org/10.1016/j.ijproman.2012.08.005.

[27] Drezner JA, Jarvaise JM, Hess RW, Hough PG, Norton D. An analysis of weapon system cost growth. 1993.

[28] Christensen DS, Antolini RC, McKinney JW. A Review of Estimate at Completion Research. J Cost Anal 1995;12:41–62. https://doi.org/10.1080/08823871.1995.10462292.

[29] McKinney JW. Estimate-At-Completion Research-A Review and Evaluation. 1991.

[30] Watkins III H. An application of Rayleigh curve theory to contract cost estimation and control. 1982.

[31] Abernethy TS. An application of the Rayleigh distribution to contract cost data. Monterey, California. Naval Postgraduate School, 1984.

[32] Gallagher MA, Lee DA. Final costs estimates for research & development programs conditioned on realized costs. Mil Oper Res J 1996;2:51–65.

https://doi.org/10.5711/morj.2.2.51.

[33] Richard W. Estimates at Completion Using Beta Curves. J Parametr 1982;15.

[34] Atkinson R, Crawford L, Ward S. Fundamental uncertainties in projects and the scope of project management. Int J Proj Manag 2006. https://doi.org/10.1016/j.ijproman.2006.09.011.

[35] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.

[36] Zeng J, An M, Smith NJ. Application of a fuzzy based decision making methodology to construction project risk assessment. Int J Proj Manag 2007;25:589–600. https://doi.org/10.1016/j.ijproman.2007.02.006.

[37] Dvir D. Transferring projects to their final users: The effect of planning and preparations for commissioning on project success. Int J Proj Manag 2005;23:257–65. https://doi.org/10.1016/j.ijproman.2004.12.003.

[38] Cheng MY, Peng HS, Wu YW, Chen TL. Estimate at completion for construction projects using evolutionary support vector machine inference model. Autom Constr 2010;19:619–29. https://doi.org/10.1016/j.autcon.2010.02.008.

[39] Wang Y-R, Yu C-Y, Chan H-H. Predicting construction cost and schedule success using artificial neural networks ensemble and support vector machines classification models. Int J Proj Manag 2012;30:470–8. https://doi.org/10.1016/j.ijproman.2011.09.002.

[40] Cheng M, Hoang N, Roy AF V, Wu Y. A novel time-depended evolutionary fuzzy SVM inference model for estimating construction project at completion. Eng Appl Artif Intell 2012;25:744–52. https://doi.org/10.1016/j.engappai.2011.09.022.

[41] Christensen DS. Project advocacy and the estimate at completion problem. J Cost Anal Manag 1996;1996:35–60.

[42] Christensen DS. Using performance indices to evaluate the estimate at completion. J Cost Anal Manag 1994;1994:17–24. https://doi.org/10.1080/08823871.1994.10462282.

[43] Narbaev T, De Marco A. An Earned Schedule-based regression model to improve cost estimate at completion. Int J Proj Manag 2014;32:1007–18. https://doi.org/10.1016/j.ijproman.2013.12.005.

[44] Narbaev T, De Marco A. Combination of Growth Model and Earned Schedule to Forecast Project Cost at Completion. J Constr Eng Manag 2014;140:04013038. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000783.

[45] Huang C-C, Cheng M-Y. Estimate at completion for construction projects Using Evolutionary Gaussian Process Inference Model. Int. Conf. Multimed. Technol., 2011, p. 4414–7. https://doi.org/10.1109/ICMT.2011.6003217.

[46] Kim E, Wells Jr WG, Duffey MR. A model for effective implementation of earned value management methodology. Int J Proj Manag 2003;21:375–82.

[47] Kim B, Reinschmidt KF. Probabilistic Forecasting of Project Duration Using Bayesian Inference and the Beta Distribution. J Constr Eng Manag 2009:178–86.

[48] Vanhoucke M. Measuring time: Improving project performance using earned value management. Springer Science \& Business Media; 2009.

[49] Haydon JJ, Reither RO. Methods of Estimating Contract Cost at Completion. ManTech Int Corp Virginia 1982;31.

[50] Covach J, Haydon JJ, Reither RO. A study to determine indicators and methods to compute estimate at completion (EAC). Virginia ManTech Int Corp 1981;30.

[51] Lollar JL. Cost Performance Analysis Program for Use on Hand-Held Programmable Calculators. 1980.

[52] Blythe AL. A Stochastic Model for Estimating Total Program Cost. ASD Reserv Report, Aeronaut Syst Div Wright-Patterson AFB, Ohio 1982.

[53] Cryer JM, Balthazor LR. Evaluation of Weighted Indices on Algorithms Utilized for Calculating Independent Estimates at Completion 1986.

[54] Totaro JA. A Logical Approach to Estimate at Completion Formulas. Progr Manag 1987;16:29–33.

[55] Olsen D, Ellsworth RW. Forecasting Techniques Employed in a Line Organization. 1976.

[56] Busse DE. A Cost Performance Forecasting Model. Air University, Maxwell AFB, Alabama, 1977.

[57] Karsch O. A cost performance forecasting concept and model. 1974.

[58] Weida WJ. A General Technique for R&D Cost Forecasting. 1977.

[59] Chacko GK. Improving Cost and Schedule Controls Through Adaptive Forecasting. Concepts J Def Syst Acquis Manag 1981;4:73–96.

[60] El-Sabban Z. Forecast of Cost/Schedule Status Utilizing Cost Performance Reports of the Cost/Schedule Control Systems Criteria: A Bayesian Approach. US Army Aviat Syst Command St Louis Missouri, AD-754576 1973.

[61] Hayes RA. An Evaluation of a Bayesian Approach to Compute Estimates at Completion for Weapon System Programs. School of Systems and Logistics, Air Force Institute of Technology (AU), 1977.

[62] Holeman Jr J. A Product Improved Method for Developing a Program Management Office Estimated Cost at Completion. Virginia: 1975.

[63] Cheng MY, Tsai HC, Hsieh WS. Web-based conceptual cost estimates for construction projects using Evolutionary Fuzzy Neural Inference Model. Autom Constr 2009;18:164–72. https://doi.org/10.1016/j.autcon.2008.07.001.

[64] Cheng MY, Roy AFV. Evolutionary fuzzy decision model for construction management using support vector machine. Expert Syst Appl 2010;37:6061–9. https://doi.org/10.1016/j.eswa.2010.02.120.

[65] Cheng M-Y, Tsai H-C, Sudjono E. Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry. Expert Syst Appl 2010;37:4224–31. https://doi.org/10.1016/j.eswa.2009.11.080.

[66] Feylizadeh MR, Hendalianpour A, Bagherpour M. A fuzzy neural network to estimate at completion costs of construction projects. Int J Ind Eng Comput 2012;3:477–84. https://doi.org/10.5267/j.ijiec.2011.11.003.

[67] Caron F, Ruggeri F, Merli A. A bayesian approach to improve estimate at completion in earned value management. Proj Manag J 2013;44:3–16. https://doi.org/10.1002/pmj.21303.

[68] Enshassi A, Mohamed S, Abushaban S. Factors affecting the performance of construction projects in the Gaza strip. J Civ Eng Manag 2009;15:269–80.

[69] Mbachu J, Nkado R. Factors constraining successful building project implementation in South Africa. Constr Manag Econ 2007;25:39–54. https://doi.org/10.1080/01446190600601297.

[70] Wiguna IPA, Scott S. Nature of the critical risk factors affecting project performance in Indonesian building contracts. 21st Annu. ARCOM Conf., vol. 1, 2005, p. 225–35.

[71] Hanson DN, Mbachu J, Nkado R. CAUSES OF CLIENT DISSATISFACTION IN THE SOUTH AFRICAN BUILDING INDUSTRY AND WAYS OF IMPROVEMENT: THE CONTRACTORS'PERSPECTIVES. Publ Master's Diss Univ Witwatersrand 2006.

[72] Mahamid I, Bruland A. Cost diviation in road construction projects: The case of Palestine. Constr Econ Build 2012;12:58–71. https://doi.org/10.5130/ajceb.v12i1.2427.

[73] Koushki PA, Al-Rashid K, Kartam N. Delays and cost increases in the construction of private residential projects in Kuwait. Constr Manag Econ 2005;23:285–94.

[74] Omran A, Abdalrahman S, Pakir AK. Project performance in Sudan construction industry: A case study. Glob J Account Econ Reseacrh 2012.

[75] Golizadeh H, Banihashemi S, Sadeghifam AN, Preece C. Automated estimation of completion time for dam projects. Int J Constr Manag 2017;17:197–209. https://doi.org/10.1080/15623599.2016.1192249.

[76] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 1943;5:115–33. https://doi.org/10.1007/BF02478259.

[77] Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychol Rev 1958. https://doi.org/10.1037/h0042519.

[78] Stewart SD, Watson G. Applications of artificial intelligence. Simulation 1985. https://doi.org/10.1177/003754978504400607.

[79] Werbos PJ. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. 1974. https://doi.org/10.1.1.41.8085.

[80] Hinton GE, Osindero S, Teh Y-W. A Fast Learning Algorithm for Deep Belief Nets. Neural Comput 2006;18:1527–54. https://doi.org/10.1162/neco.2006.18.7.1527.

[81] Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE. A survey of deep neural

network architectures and their applications. Neurocomputing 2017;234:11–26. https://doi.org/10.1016/j.neucom.2016.12.038.

[82] Lu P, Chen S, Zheng Y. Artificial intelligence in civil engineering. Math Probl Eng 2012;2012. https://doi.org/10.1155/2012/145974.

[83] Baştanlar Y, Özuysal M. Introduction to machine learning. Methods Mol Biol 2014. https://doi.org/10.1007/978-1-62703-748-8_7.

[84] Kotsiantis SB. Supervised machine learning: A review of classification techniques. Inform 2007. https://doi.org/10.1115/1.1559160.

[85] Deng L, Yu D. Deep learning methods and apllications. Bmj 1999. https://doi.org/10.1136/bmj.319.7209.0a.

[86] Bengio Y. Deep Learning of Representations for Unsupervised and Transfer Learning. J Mach Learn Res 2012. https://doi.org/10.1109/IJCNN.2011.6033302.

[87] Schmidhuber J. Deep Learning in neural networks: An overview. Neural Networks 2015;61:85–117. https://doi.org/10.1016/j.neunet.2014.09.003.

[88] Bengio Y. Learning Deep Architectures for AI. Found Trends® Mach Learn 2009;2:1–127. https://doi.org/10.1561/2200000006.

[89] Ngiam J, Khosla A, Kim M. Multimodal deep learning. … Mach Learn (… 2011. https://doi.org/10.1145/2647868.2654931.

[90] Introduction a S. Neural Networks. Neural Networks 1996;7:509. https://doi.org/10.1016/0893-6080(94)90051-5.

[91] Tadeusiewicz R. Neural networks: A comprehensive foundation. Control Eng Pract 1995;3:746–7. https://doi.org/10.1016/0967-0661(95)90080-2.

[92] Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci 1982;79:2554–8. https://doi.org/10.1073/pnas.79.8.2554.

[93] Cawley GC, Talbot NL. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. J Mach Learn Res 2010. https://doi.org/10.1016/j.biopha.2003.08.027.

[94] Arulkumaran K, Deisenroth MP, Brundage M, Bharath AA. Deep reinforcement learning: A brief survey. IEEE Signal Process Mag 2017. https://doi.org/10.1109/MSP.2017.2743240.

[95] Deng L, Yu D. Deep learning: methods and applications. Found Trends Signal Process 2014;7:197–387.

[96] Vapnik V. The Nature of statistical Learning Theory. 1995.

[97] Lingras P, Butz CJ. Rough support vector regression. Eur J Oper Res 2010;206:445–55. https://doi.org/10.1016/j.ejor.2009.10.023.

[98] Yaseen ZM, Kisi O, Demir V. Enhancing long-term streamflow forecasting and predicting using periodicity data component: application of artificial intelligence. Water Resour Manag 2016;30:4125–51.

[99] Raghavendra S, Deka PC. Support vector machine applications in the field of

hydrology: A review. Appl Soft Comput J 2014;19:372–86. https://doi.org/10.1016/j.asoc.2014.02.002.

[100] Platt J. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Adv Large Margin Classif 1999;10:61–74. https://doi.org/10.1.1.41.1639.

[101] Holland JH. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT press; 1992.

[102] Iba H, Aranha CC. Introduction to genetic algorithms. Adapt Learn Optim 2012. https://doi.org/10.1007/978-3-642-27648-4_1.

[103] Harik GR, Lobo FG, Goldberg DE. The compact genetic algorithm. IEEE Trans Evol Comput 1999. https://doi.org/10.1109/4235.797971.

[104] Yang J, Honavar V. Feature subset selection using genetic algorithm. IEEE Intell Syst Their Appl 1998. https://doi.org/10.1109/5254.671091.

[105] Osborne AR. Simple, Brute-force computation of theta functions and beyond. Int Geophys 2010;97:489–99. https://doi.org/10.1016/S0074-6142(10)97020-4.

[106] Heule MJH, Kullmann O. The Science of Brute Force. Commun ACM 2017;60:70–9. https://doi.org/10.1145/3107239.

[107] Hofmann M, Klinkenberg R. RapidMiner: Data mining use cases and business analytics applications. CRC Press; 2016.

[108] Ristoski P, Bizer C, Paulheim H. Mining the web of linked data with rapidminer. J Web Semant 2015;35:142–51.

[109] Amer M, Goldstein M. Nearest-neighbor and clustering based anomaly detection algorithms for rapidminer. Proc. 3rd RapidMiner Community Meet. Conf. (RCOMM 2012), 2012, p. 1–12.

[110] Jungermann F. Information extraction with rapidminer. Proc. GSCL Symp. und eHumanities, Citeseer; 2009, p. 50–61.

[111] Bayram S. Duration prediction models for construction projects: In terms of cost or physical characteristics? KSCE J Civ Eng 2017;21:2049–60.

[112] Kim Y-J, Yeom D-J, Kim YS. Development of construction duration prediction model for project planning phase of mixed-use buildings. J Asian Archit Build Eng 2019;18:586–98.

[113] Kaka A, Price ADF. Relationship between value and duration of construction projects. Constr Manag Econ 1991;9:383–400.

[114] Chen L-H, Raitio T, Valentini-Botinhao C, Yamagishi J, Ling Z-H. DNN-based stochastic postfilter for HMM-based speech synthesis. INTERSPEECH, 2014, p. 1954–8.

[115] Sivapatham S, Ramadoss R, Kar A, Majhi B. Monaural speech separation using GA-DNN integration scheme. Appl Acoust 2020;160:107140.

[116] Cheng MY, Hoang ND, Roy AFV, Wu YW. A novel time-depended evolutionary fuzzy SVM inference model for estimating construction project at completion.

Eng Appl Artif Intell 2012;25:744–52. https://doi.org/10.1016/j.engappai.2011.09.022.

[117] Nassar N, AbouRizk S. Practical application for integrated performance measurement of construction projects. J Manag Eng 2014;30:4014027.

[118] Barraza GA, Back WE, Mata F. Probabilistic forecasting of project performance using stochastic S curves. J Constr Eng Manag 2004;130:25–32.

[119] Juszczyk M. The challenges of nonparametric cost estimation of construction works with the use of artificial intelligence tools. Procedia Eng 2017;196:415–22.

[120] Va V, Shimizu T, Bansal G, Heath Jr RW. Millimeter wave vehicular communications: A survey. Found Trends® Netw 2016;10:1–118.

[121] Teyeb O, Kazmi M, Mildh G. Mobility state aware mobile relay operation 2017.

[122] Lázaro ODM, Mohammed WM, Ferrer BR, Bejarano R, Lastra JLM. An Approach for adapting a Cobot Workstation to Human Operator within a Deep Learning Camera. 2019 IEEE 17th Int. Conf. Ind. Informatics, vol. 1, IEEE; 2019, p. 789–94.

[123] Liu B, Wei Y, Zhang Y, Yang Q. Deep Neural Networks for High Dimension, Low Sample Size Data. IJCAI, 2017, p. 2287–93.

[124] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. Int J Comput Vis 2015;115:211–52.

[125] Kloock J, Schiller U. Marginal costing: cost budgeting and cost variance analysis. Manag Account Res 1997;8:299–323.

[126] Fleming QW, Koppelman JM. Earned value project management, Project Management Institute; 2016.

[127] Keil M, Truex DP, Mixon R. The effects of sunk cost and project completion on information technology project escalation. IEEE Trans Eng Manag 1995;42:372–81.

[128] Lipke W. Schedule is different. Meas News 2003;31:31–4.

[129] Lipke W, Zwikael O, Henderson K, Anbari F. Prediction of project outcome. The application of statistical methods to earned value management and earned schedule performance indexes. Int J Proj Manag 2009;27:400–7. https://doi.org/10.1016/j.ijproman.2008.02.009.

[130] Short JW. Using schedule variance as the only measure of schedule performance. Cost Eng 1993;35:35.

[131] Christensen DS, Heise SR. Cost performance index stability. Natl Contract Manag J 1993;25:7–15.

[132] Christensen D, Payne K. Cost performance index stability: fact or fiction? J Parametr 1992;12:27–40.

[133] Payne KI. An investigation of the stability of the cost performance index. AIR FORCE INST OF TECH WRIGHT-PATTERSONAFB OH; 1990.

[134] Kim S. Project success indicators focusing on residential projects: Are

schedule performance index and cost performance index accurate measures in earned value? Can J Civ Eng 2009;36:1700–10.

[135] Petter JL. An analysis of stability properties in earned value management's cost performance index and earned schedule's schedule performance index 2014.

[136] Nassar KM, Gunnarsson HG, Hegab MY. Using Weibull analysis for evaluation of cost and schedule performance. J Constr Eng Manag 2005;131:1257–62.

[137] Camilleri MJT, Jaques RA, Isaacs NP. Climate change impacts on building performance. BRANZ; 2001.

[138] Camilleri MJT. Implications of climate change for the construction sector: houses. BRANZ; 2000.

[139] El-Sawalhi N, Mahdi M. Influence of Climate Change on the Lifecycle of Construction Projects at Gaza Strip. J Constr Eng Proj Manag 2015;5:1–10.

[140] Elfahham Y. Estimation and prediction of construction cost index using neural networks, time series, and regression. Alexandria Eng J 2019;58:499–506.

[141] Yaseen ZM, Tran MT, Kim S, Bakhshpoori T, Deo RC. Shear strength prediction of steel fiber reinforced concrete beam using hybrid intelligence models: A new approach. Eng Struct 2018;177:244–55. https://doi.org/10.1016/j.engstruct.2018.09.074.

[142] Williams TP. Predicting changes in construction cost indexes using neural networks. J Constr Eng Manag 1994;120:306–20.

[143] Abdullah SS, Abdul Malek M, Mustapha A, Aryanfar A. Hybrid of Artificial Neural Network-Genetic Algorithm for Prediction of Reference Evapotranspiration (ET0) in Arid and Semiarid Regions. J Agric Sci 2014;6:191–200. https://doi.org/10.5539/jas.v6n3p191.

[144] Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. Geosci Model Dev 2014;7:1247–50. https://doi.org/10.5194/gmd-7-1247-2014.

[145] Nash JE, Sutcliffe J V. River flow forecasting through conceptual models part I - A discussion of principles. J Hydrol 1970;10:282–90. https://doi.org/10.1016/0022-1694(70)90255-6.

[146] Willmott CJ. On the validation of models. Phys Geogr 1981;2:184–94.

[147] Zeng J, Qiao W. Short-term solar power prediction using a support vector machine. Renew Energy 2013;52:118–27. https://doi.org/10.1016/j.renene.2012.10.009.

[148] Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 2005;30:79–82.

[149] Willmott CJ, Matsuura K, Robeson SM. Ambiguities inherent in sums-of-squares-based error statistics. Atmos Environ 2009;43:749–52.

[150] Puniya M, Singh RB. Correlation and Regression Analysis n.d.

Appendix A The applied historical data set of the construction projects

Projects	CV	SV	CPI	SPI	Subcontractor billed index	Owner billed index	Change order index	CCI	Climate effect index	EAC
1-1	2400	2200	4	3.2	1	1	1	1	1	7975
1-2	5710	4610	3.114815	2.213158	1	1	1	1	1	10241.38
1-3	8750	6850	2.62037	1.938356	1.33	0.75	1	1	0.99	12173.85
1-4	8541.111	5241.111	2.041599	1.455749	1.67	0.75	1	1	0.97	15625.01
1-5	10720	7820	1.864516	1.511111	1.52	0.75	1	1	0.62	17109
1-6	6790	4490	1.399412	1.232642	1	0.68	1	1.12	0.63	22795.29
1-7	-7950	-11250	0.6025	0.517167	1	0.64	1	1.13	0.82	52946.06
1-8	-3845	-6745	0.827578	0.732341	0.99	0.53	1	1.13	0.74	38546.19
1-9	-7780	-10880	0.668936	0.590977	1	0.42	1	1.13	0.62	47687.66
2-1	-20386.7	-24186.7	0.177957	0.154312	0.67	1.02	1	1.14	0.65	179256.8
2-2	-20880	-25080	0.209091	0.180392	0.75	1.52	1	1.15	1	152565.2
2-3	-18595	-22295	0.340603	0.301097	1.22	0.77	1	1.12	0.98	93657.47
2-4	3450	2950	4.45	2.966667	1.22	1.02	1	1.11	0.99	2898.876
2-5	2500	1450	2.470588	1.527273	1.08	1.08	1	1.12	0.96	5221.429
2-6	4712.5	3362.5	2.625	1.791176	1.09	0.88	1	1.13	0.96	4914.286
2-7	3844.444	2194.444	1.739316	1.320357	0.81	1	1	1.18	0.72	7416.708
2-8	-165	-2615	0.976087	0.720321	0.82	1	1	1.11	0.62	13216.04
2-9	-3700	-6250	0.493151	0.365482	0.82	1	1	1.12	0.62	26158.33
2-10	-4300	-7000	0.455696	0.339623	0.82	1	1	1.12	0.62	28308.33
2-11	-4565	-7515	0.469186	0.349351	1.57	1.02	1	1	0.62	27494.42
2-12	-6375	-9425	0.287709	0.214583	1.57	1.05	1	1	1	44836.89
3-1	-6915	-10115	0.2524	0.1875	1.62	0.62	1	1	1	51102.
3-2	-8300	-11500	0.126316	0.094488	1.62	0.51	1	1	1	102125
3-3	-8500	-11700	0.123711	0.093023	1.75	0.48	1	1	0.82	104275
3-4	1550	1450	16.5	8.25	1.75	1.08	1	1.11	0.92	1154.545
3-5	1462.5	1062.5	3.4375	2.0625	1	1	1	1.12	0.92	5541.818
3-6	3662.5	2762.5	3.616071	2.201087	1	1	1	1.19	0.92	5268.148
3-7	8120	7070	3.706667	2.745679	1	0.93	1	1.16	0.74	5139.388
3-8	5652.5	4552.5	1.890157	1.611074	1	0.87	1	1	0.68	10078.53

3-9	585	-1065	1.073125	0.889637	0.54	0.84	1	1	0.82	17751.89
3-10	4797.5	3097.5	1.492051	1.270524	0.64	0.82	1	1	0.67	12767.66
4-1	-1385	-3235	0.878509	0.755849	0.64	0.52	1	1	0.67	21684.47
4-2	-2515	-3915	0.814391	0.738127	1.23	0.48	1	1	0.82	23391.71
4-3	-10360	-12110	0.270423	0.240752	1.47	0.67	1.02	1.13	0.92	70445.31
4-4	-10862.5	-13112.5	0.299194	0.261268	1.47	1.02	1.02	1.13	0.92	63671.16
4-5	-11762.5	-14412.5	0.282774	0.243438	1	1.08	1.02	1.13	1	67368.19
4-6	3075	2975	11.25	8.4375	1	1.06	1.01	1.15	1	1693.333
4-7	3420.833	3070.833	3.138021	2.574786	1	1.08	1.01	1.14	1	6070.705
4-8	4325	3725	2.601852	2.128788	1.52	0.67	1.01	1.11	0.92	7321.708
4-9	3355	2205	1.79881	1.41215	1.52	0.84	1.01	1.11	0.82	10590.34
4-10	4875	3825	1.716912	1.487261	0.99	0.83	1.01	1.12	0.71	11095.5
4-11	-2320	-3670	0.690667	0.585311	0.97	1.22	1.01	1.12	0.72	27582.05
4-12	-4500	-5450	0.476744	0.429319	0.97	1.42	1.01	1.12	0.72	39958.54
4-13	-5161.4	-6611.4	0.459536	0.398961	0.61	1.53	1.01	1.13	0.73	41454.83
5-1	-7100	-8850	0.400844	0.349265	0.72	1.84	1.01	1.13	0.75	47524.74
5-2	-8450	-10700	0.342412	0.291391	0.72	1	1.01	1.13	0.62	55634.66
5-3	-8933.33	-11583.3	0.398429	0.338095	0.75	1	1.01	1.11	0.61	47812.82
5-4	-9710	-12310	0.409726	0.353806	1.42	1	1.01	1.08	0.84	46494.44
5-5	800	700	9	4.5	1.42	1.02	1.01	1.18	0.84	1422.222
5-6	1425	1125	4.5625	2.607143	0.61	1.06	1.01	1.02	0.84	2805.479
5-7	1325	375	2.472222	1.202703	0.61	0.63	1.01	1.11	0.72	5177.528
5-8	1682.5	232.5	1.80119	1.065493	1.63	0.67	1.01	1.13	0.72	7106.411
5-9	487.5	-1212.5	1.174107	0.730556	1.63	0.45	1.01	1.14	0.64	10901.9
5-10	-262.143	-2462.14	0.929151	0.582688	1	0.72	1.01	1.12	0.63	13776.02
6-1	-950	-3500	0.776471	0.485294	1	0.71	1.01	1.12	0.63	16484.85
6-2	-3260	-6110	0.259091	0.157241	1	0.64	1.01	1	0.63	49403.51
6-3	-1840	-5040	0.642718	0.396407	1	0.82	1.01	1	1	19915.41
6-4	-1390	-4790	0.779365	0.506186	0.82	0.92	1.01	1	1	16423.63
6-5	-4100	-8200	0.426573	0.271111	0.87	1.08	1.03	1	0.95	30006.56
6-6	-4100	-8600	0.506024	0.328125	0.92	1.07	1.02	1	0.95	25295.24
6-7	1333.333	833.3333	2.333333	1.555556	0.92	1	1.02	1	0.95	11700
6-8	1500	1000	1.75	1.4	1.02	1	1.02	1	0.92	15600
6-9	3750	3400	1.974026	1.809524	1.32	1	1.02	1	0.81	13829.61
6-10	1150	150	1.239583	1.025862	1.32	1	1.03	1	0.81	22023.53
6-11	3010	410	1.418056	1.041837	1	0.99	1.03	1.13	0.71	19251.71
6-12	1420	-1880	1.142	0.858647	1	1.05	1.03	1.12	0.74	23905.43
7-1	-3720	-7720	0.684746	0.511392	1	0.42	1.02	1.15	0.74	39868.81

7-2	-1500	-6200	0.89726	0.678756	1.22	0.71	1.01	1.02	0.74	30425.95
7-3	-7400	-12500	0.54321	0.413146	1.22	0.63	0.96	1.13	0.84	50256.82
7-4	-12266.7	-18066.7	0.318519	0.240896	1.44	0.72	1.02	1.13	0.84	85709.3
7-5	-12900	-18900	0.364532	0.281369	1.44	0.52	1.01	1.13	0.62	74890.54
7-6	-18300	-24300	0.140845	0.10989	1.44	0.71	0.94	1.13	0.63	193830
7-7	125	-125	1.5	0.75	1	0.82	0.94	1.12	0.94	15366.67
7-8	-50	-450	0.857143	0.4	1	0.63	0.94	1	0.96	26891.67
7-9	2450	1950	4.266667	2.56	0.83	0.91	1.02	1.12	0.96	5402.344
7-10	2890	1990	2.229787	1.612308	0.83	0.72	1.02	1	1	10337.31
7-11	1966.667	366.6667	1.441948	1.060606	0.83	1.08	0.94	1.12	1	15985.32
8-1	2075	-325	1.340164	0.961765	0.62	1.22	0.91	1.05	0.82	17199.39
8-2	-1350	-4300	0.826923	0.6	0.62	1.33	0.94	1.07	0.84	27874.42
8-3	-3840	-7090	0.595789	0.443922	0.62	1.07	1.01	1.11	0.82	38688.16
8-4	-6300	-9850	0.461538	0.354098	1.71	0.63	1.01	1	0.72	49941.67
8-5	-7400	-11450	0.478873	0.372603	1.72	0.42	1.01	1	0.62	48133.82
8-6	-12300	-17350	0.190789	0.14321	0.99	1.33	1.01	1	0.71	120813.8
8-7	-12775	-18325	0.27	0.204989	0.67	1.33	1.01	1	0.71	85370.37
8-8	2333.333	1833.333	3.333333	2.222222	0.67	1	1.01	1	0.92	10650
8-9	2200	1500	2.222222	1.6	0.82	1	1.01	1	0.92	15975
9-1	1783.333	83.33333	1.540404	1.016667	0.82	1	1.01	1	0.92	23045.9
9-2	3550	950	1.657407	1.11875	1.33	1	1.01	1.15	0.64	21418.99
9-3	-400	-4300	0.953488	0.656	1.33	1	1.01	1.16	0.77	37231.71
9-4	-4933.33	-9633.33	0.54321	0.378495	1.33	0.99	1.01	1.12	0.83	65352.27
9-5	-6733.33	-12133.3	0.465608	0.325926	0.71	0.52	1.01	1.12	1	76244.32
9-6	-8433.33	-15333.3	0.441501	0.30303	0.71	0.42	1.01	1.11	0.93	80407.5
9-7	-9200	-17800	0.5	0.340741	1.54	0.61	1.01	1.18	0.93	71000
9-8	-13566.7	-23666.7	0.350877	0.236559	1.54	0.82	1.01	1.02	0.93	101175
9-9	-17066.7	-28166.7	0.238095	0.159204	1.54	0.72	1.01	1.12	0.64	149100
9-10	-20150	-31750	0.156904	0.105634	1.22	0.92	1.01	1.13	0.72	226253.3
9-11	3525	2725	4.525	2.513889	1.22	0.43	1.01	1.13	0.81	4972.376
10-1	3840	2840	2.476923	1.788889	0.73	0.52	1.01	1.15	0.94	9083.851
10-2	266.6667	-1233.33	1.074074	0.75817	0.73	0.58	0.91	1.14	0.61	20948.28
10-3	3440	1440	1.614286	1.189474	0.82	0.63	0.91	1.13	0.68	13938.05
10-4	2516.667	-383.333	1.296078	0.966374	0.82	1.02	0.94	1.11	0.77	17360.06
10-5	-2395.83	-6295.83	0.778164	0.571712	1.02	1.08	1.02	1.14	0.74	28914.23
10-6	-470	-3970	0.965693	0.769186	1.02	1.44	1.03	1.13	0.62	23299.32
10-7	-8620	-12020	0.436601	0.357219	1.08	1.55	0.92	1.02	0.63	51534.43
10-8	-9980	-13880	0.405952	0.329469	1.08	1.53	0.92	1.15	0.65	55425.22

10-9	-14800	-18700	0.159091	0.130233	0.83	1	1.01	1.11	0.65	141428.6
10-10	-15300	-19200	0.154696	0.127273	0.83	1	0.92	1	1	145446.4
10-11	-16950	-21100	0.076294	0.062222	1.57	1	1.05	1	0.84	294910.7
10-12	7100	6750	7.1739	5.5	1.57	1	0.96	1	0.82	3923.9
10-13	2100	1000	1.875	1.285714	1.57	1	0.96	1	0.92	15013.33
11-1	6125	4675	2.512346	1.85	1	1.44	0.96	1	0.92	11204.67
11-2	4740	2840	1.623684	1.298947	1	1.82	0.96	1	1	17337.12
11-3	3833.333	2333.333	1.294872	1.16092	1	1.63	0.96	1.13	1	21739.6
11-4	-7508.75	-9108.75	0.489201	0.441181	1	1.23	0.96	1.12	0.72	57542.85
11-5	-7616.67	-9216.67	0.564762	0.517452	1.3	1.44	0.96	1	0.72	49844.01
11-6	-14291.7	-16491.7	0.245822	0.220252	1.64	1.44	0.96	1.15	0.63	114513.6
11-7	-14210	-16810	0.355556	0.318053	1.64	1	0.96	1	0.64	79171.88
11-8	-18850	-21550	0.160356	0.143141	1.64	1	0.96	1.13	1	175546.5
11-9	-20450	-23650	0.127932	0.11257	1	1	0.96	1.14	0.92	220039.2
11-10	-20450	-23650	0.180361	0.159858	1	1	0.96	1.12	0.92	156076.1
11-11	850	650	5.25	2.625	1.84	1.03	0.96	1.11	0.93	2161.905
12-1	725	325	2.45	1.361111	1.84	1.05	0.96	1.12	1	4632.653
12-2	1800	600	2.44	1.244898	1.84	1.06	0.96	1.05	0.82	4651.639
12-3	1733.75	-466.25	1.753804	0.896389	1.84	1.06	1.02	1.09	0.82	6471.645
12-4	665	-2085	1.218033	0.640517	1.84	0.43	1.05	1	0.82	9318.304
12-5	-272.5	-3772.5	0.935119	0.510065	1	0.52	1.02	1.13	0.75	12137.49
12-6	-2120	-5870	0.548936	0.305325	1	0.61	1.05	1.14	0.75	20676.36
12-7	-1812.5	-5762.5	0.664352	0.38369	1	0.61	1.02	1.14	0.75	17084.32
12-8	-4000	-8150	0.354839	0.21256	1	0.75	106	1.08	0.68	31986.36
12-9	-4800	-9150	0.314286	0.193833	1	0.75	1.02	1.08	0.64	36113.64
12-10	1275	1075	3.125	2.3437	1	0.82	1.02	1.13	0.64	4672
12-11	4366.667	4216.667	4.493333	4.011905	1.22	0.82	104	1.12	0.64	3249.258
12-12	3390	3340	2.189474	2.151724	0.67	1.04	0.92	1.11	0.82	6668.269
13-1	5858.333	6458.333	2.018841	2.254045	0.67	1.04	1.02	1	0.82	7231.874
13-2	1400.833	2300.833	1.167764	1.308837	1.05	1.44	1.01	1.13	1	12502.52
13-3	-1317.5	-267.5	0.875708	0.97199	1.05	1.44	1	1.12	1	16672.23
13-4	-4475	-3525	0.61588	0.670561	0.82	1.33	1	1.11	0.83	23705.92
13-5	-10275	-9225	0.154321	0.168919	0.82	0.63	1	1	0.83	94608
13-6	-11350	-10400	0.095618	0.103448	1.62	0.62	1	1.04	0.94	152691.7
13-7	-12150	-11300	0.129032	0.137405	1.62	0.82	1	1.13	0.94	113150
13-8	-13950	-13100	0.097087	0.10274	1.62	0.82	1	1.12	0.94	150380
13-9	2666.667	2166.667	3.666667	2.444444	0.74	1.77	1	1.12	1	10731.82
13-10	8893.333	8693.333	3.505164	3.318222	0.74	1.42	1	1.12	1	11226.29

13-11	5329.167	4729.167	1.895658	1.72201	1.53	1.43	1	1.05	1	20757.96
13-12	2063.333	1163.333	1.225501	1.115755	1.53	0.62	1	1.05	0.75	32109.32
13-13	-337.5	-1037.5	0.972222	0.919261	0.62	0.55	1	1.14	0.76	40474.29
13-14	-4803.33	-5003.33	0.70258	0.693986	0.74	0.81	1	1.14	1	56007.86
14-1	-10000	-9500	0.469496	0.482289	0.74	0.81	1	1.14	1	83813.28
14-2	-8450	-7750	0.633406	0.653244	0.99	1	1	1.14	1	62124.49
14-3	-10650	-10750	0.616216	0.614004	0.99	1	1	1.08	0.82	63857.46
14-4	-15883.3	-16683.3	0.496566	0.484286	0.99	1	1	1	0.84	79244.2
14-5	-22500	-24000	0.354376	0.339752	1.44	1	1	1	0.84	111040.3
14-6	-23450	-25150	0.377158	0.360864	1.52	1	1	1	0.85	104332.9
14-7	4720	4420	7.742857	5.42	1.52	1.05	1	1	1	2983.395
14-8	8180	8180	5.09	5.09	1.08	1.05	1	1.13	0.77	4538.31
14-9	4074.286	4074.286	1.970068	1.970068	1.08	1.05	1	1.01	0.83	11725.48
14-10	350	550	1.056452	1.091667	0.74	1.33	1	1.01	0.64	21865.65
14-11	1832.143	2332.143	1.229018	1.310952	0.74	1.33	1	1.07	0.64	18795.5
14-12	1490.476	1990.476	1.135498	1.189569	0.74	1.22	1	1.24	1	20343.5
14-13	-5050	-4850	0.636691	0.645985	1.44	1.47	1	1.18	0.95	36281.36
15-1	-4650	-4550	0.734286	0.738506	1.44	1.47	1	1.02	0.93	31459.14
15-2	-14500	-14600	0.263959	0.262626	1.66	1.47	1	1.12	0.72	87513.46
15-3	-16366.7	-16666.7	0.249235	0.245852	1.66	1.34	1	1.15	0.84	92683.44
15-4	-20800	-21100	0.087719	0.08658	1.66	1.34	1	1.16	0.84	263340
15-5	520	500	7.5	6	0.71	1.33	0.92	1.02	0.83	1246.667
15-6	1250	1200	3.777778	3.4	0.71	0.68	1.05	1.05	0.64	2475
15-7	712.5	612.5	1.838235	1.644737	1.06	0.61	0.94	1.14	0.72	5086.4
15-8	1075	725	1.632353	1.353659	1.06	0.61	1.08	1.13	0.72	5727.928
15-9	1603.333	1003.333	1.562573	1.290821	1.44	0.61	0.92	1.11	0.82	5983.72
15-10	1162.5	562.5	1.261236	1.111386	1.44	1.02	1.01	1	0.94	7413.363
15-11	-1216.67	-1766.67	0.788406	0.719577	1.33	1.05	1	1.12	0.93	11859.38
15-12	-3136.67	-3886.67	0.548681	0.495238	1.33	1.05	1.05	1.12	0.72	17040.87
15-13	-4395	-5395	0.457407	0.407143	1.23	1.05	1	1.13	1	20441.3
15-14	-7600	-8600	0.08982	0.080214	1.23	1.08	1	1.13	1	104096.7

CURRICULUM VITAE

PERSONAL INFORMATION

Name Surname : Karrar Raoof Kareem KAMOONA

Date of birth and place : 29.01.1981

Foreign Languages : English, Arabic

E-mail : Karrar19811940@gmail.com

Phone Number : 00905456161939

EDUCATION

Degree	**Department**	**University**	**Date of Graduation**
Master Degree	Construction Engineering	SHUATS	2013-2014
Undergraduate	Civil Engineering	Kufa University	2004-2005
High School	Baccalaureate	AL-Mufeed School	1999-2000

WORK EXPERIENCE

Year	**Corporation/Institute**	**Enrollment**
2005	Ministry of Electricity Of Iraq	Engineer

PUBLICATIONS FROM THE THESIS

Contact Information: Karrar19811940@gmail.com

Papers

1. Implementation of Genetic Algorithm Integrated with the Deep Neural Network for Estimating at Completion Simulation. 2019

2. Study of Management and Control of Waste Construction Materials in Civil Construction Project. 2013