



ISSN: 2958-8995. 2958-8987

Doi: 10.59799/APPP6605

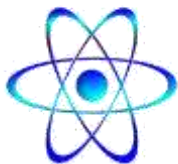
No: 8 Val:1/ January / 2025

A special number of the facts of the Mathematics, Computer and
International Statistics Conference

Journal of Natural and Applied Sciences **URAL**

A Quarterly Multidisciplinary Scientific Journal Issued by European
Academy for Development and Research / Brussels and Center of
Research and Human Resources Development Ramah- Jordan

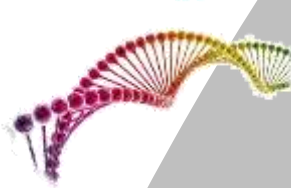
PHYSICS



Chemistry



Biology



MATHEMATICS



Pharmacy



Engineering



Medicine



Veterinary Medicine



Geology



Dentistry



1 computer



Agriculture



Editorial Team			
Prof. Dr. Ghassan Ezzulddin Arif	Tikrit University\ College of Education for Pure Science's\ Department of Mathematics.	Iraq	Editor-in-Chief of the Journal
Assist. Prof. Baraa Mohammed Ibrahim Al-Hilali	University of Samarra\ College of Education\ Biology Department	Iraq	Managing Editor of the Journal
Asst. inst. Alyaa Hussein Ashour	University of Mashreq/ College of Medical Sciences Technologies Department of Medical Physics	Iraq	Editorial Secretary of the Journal

Prof. Dr. Younis A. Rasheed	Al-Iraqia University, College of Medicine	Iraq
Assist. Prof. Dr. Hadeer Akram Al-Ani	Dept. of Public Health Sciences UC Davis School of Medicine	USA
Assist. Prof. Dr. Jawdat Akeel Mohammad Alebraheem	College of Science Al-Zulfi Majmaah University, Al-Majmaah	KSA
Assist. Prof. Dr. Almbrok Hussin Alsonosi OMAR	Sebha University	Libya
Assist. Prof. Dr. Saad Sabbar Dahham	University of Technology and Applied Sciences	Sultanate oman

Advisory and Scientific Board			
Prof. Dr. Ahamed Saied Othman	Tikrit University	Iraq	Head
Prof. Dr. Salih Hamza Abbas	University of Basrah	Iraq	Member
Prof. Dr. Leith A. Majed	University of Diyala	Iraq	Member
Assist. Prof. Dr Ali Fareed Jameel	Institute of Strategic Industrial Decision Modeling (ISIDM), School of Quantitative Sciences (SQS), University Utara (UUM), 06010 Sintok	Malaysia	Member
Assist. Prof. Mustafa Abdullah Theyab	University of Samarra	Iraq	Member
Dr. Modhi Lafta Mutar	The Open Educational College, Iraqi Ministry of Education, Thi-Qar	Iraq	Member
Dr. Asaad Shakir Hameed	Quality Assurance and Academic Performance Unit, Mazaya University College, Thi-Qar, Iraq.	Iraq	Member

Ahmad Mahdi Salih Alaubaydi	Assist. Lect.; PhD Student in the University of Sciences USM, Malaysia	Malaysia	Member
Assist. Prof. Dr. Qutaiba Hommadi Mahmood Al.Samarrraie	University of Samarra/College of Applied Sciences/ Department of Biotechnology	Iraq	Member
Ph.D. Ali Mahmood Khalaf	Gujarat University	India	Member
Dr. Amel D. Hussein	Wasit University	Iraq	Member

Focus & Scope:

Journal of Natural and Applied Sciences URAL

Journal welcomes high quality contributions investigating topics in the fields of Biology, physics, computer science, Engineering, chemistry, Geology, Agriculture, Medicine, Mathematics, Pharmacy, Veterinary, Nursing, Dentistry, and Environment.

Publication specializations in the journal	
Biology	Chemistry
Physics	Geology
Computer	Agriculture
Engineering	Mathematics
Medicine	Pharmacy
Veterinary	Dentistry Veternity,
Environment	Nursing

The Journal is published in English and Arabic

General Supervisor of the Journal

Prof. Dr. Khalid Ragheb Ahmed Al-Khatib

Head of the Center for Research and Human

Resources Development Ramah – Jordan

Managing Director:

Dr. Mosaddaq Ameen Ateah AL – Doori

Linguistic Reviewer Team

Prof. Dr. Lamiaa Ahmed Rasheed

Tikrit University/College of Education for Women

Asst. Prof. Ahmed Khalid Hasoon

Tikrit University/ College of Education for Women

Asst. Prof. Dr. Mohammad Burjess

Tikrit University/ College of Education

Administrative Title of the Journal:

Amman\ Jordan\ Wasfi Al-Tal \ Gardens

Phone: +962799424774

Index			
No.	Research Title	Researcher	Page No.
1.	Integration of Machine Learning with CFD for Enhanced Fluid Flow Predictions	Bourair Al-Attar	6- 28
2.	Exploring the Intersection of Cloud Computing and Blockchain Technologies: Use Cases in Decentralized Applications, Supply Chain Tracking, and Digital Identity	Ahmed Nafea Ayesh	29-50
3.	A new Survey on Metaheuristic Algorithms for Solving the Optimization Problems	Assistant Prof. Samera Khaleel Ibrahim Assistant Prof. Dr. Zainib Hatif Abbas	51-64
4.	Data Security Algorithm Using Magic Square With Order Five Based on GF(28) Numbers	Ibrahim Malik ALattar ¹ , Abdul Monem S. Rahma ²	65-79
5.	Estimation of the reliability using Lognormal Distribution	Rabab Abdulrida Saleh (1) Iqbal Mahmood alwan (2)	80-88
6.	Estimation Of Poisson Regression Model Using Robust Mallows Quasi-Maximum likelihood Method	A.Prof.Dr.Iqbal Mahmoud Alwan ¹ and Asaour Ammar Ahmed The Mashhadani ²	89-106
7.	الذكاء الشخصي لدى طلبة قسم الرياضيات وعلاقته بالبخل المعرفي لديهم	أ.د. عبدالواحد حميد الكبيسي أ.م.د. فائدة ياسين البديري	107-116
8.	Estimation of the multiple linear regression model for the production quantity of Al-Zahi in Al-Mamoun Factory using the (r-(k-d)) method in the presence of the problems of autocorrelation and multicollinearity together	Saja Mohammad Hussein Zainab abd alsatar	117-127
9.	A Partial Maximum Likelihood Method to Estimate Cox Model for Censoring Data From the Right with Application	Ibrahim Khalil Hussein Entsar Arba Fadam	128-149

10	Study of Chemical and Health Problems by Kuffi Issa Transform	Ahmad Issa ¹ , Emad A. Kuffi ²	151-160
11	العلاقة بين علوم الرياضيات ، الإحصاء والحاسبات Mathematics, Statistics & Computer Science (MSCS)	الأستاذ الدكتور فاضل عبدالعباس العابدي	161-171

Integration of Machine Learning with CFD for Enhanced Fluid Flow Predictio

1st Bourair Al-Attar
College of Medicine University of Al-Ameed
Karbala PO Box 198, Iraq

Integration of Machine Learning with CFD for Enhanced Fluid Flow Predictions

1st Bourair Al-Attar
College of Medicine
University of Al-Ameed
Karbala PO Box 198, Iraq
line 5: email address or ORCID

Abstract—this paper addresses the rising demand for accurate and efficient fluid flow predictions in computational fluid dynamics (CFD), which typically confront restrictions due to complicated geometries and turbulence models. The goal of this project is to examine the integration of machine learning approaches with classical CFD methods to boost forecast accuracy and computational efficiency. Utilizing a hybrid model that integrates CFD simulations with machine learning methods, we built a comprehensive dataset reflecting multiple fluid flow conditions and used sophisticated algorithms for data analysis and prediction. Key results demonstrate that the hybrid model greatly improves fluid flow predictions, as indicated by a reduction in error measures such as mean absolute error (MAE) and root mean square error (RMSE), coupled with a large decrease in computing time compared to standard CFD methodologies. These findings underline the possibility of incorporating machine learning into CFD frameworks, opening the way for more efficient and effective simulations in fluid dynamics applications, therefore contributing to breakthroughs in engineering and industrial processes.

Keywords—computational fluid dynamics, machine learning, hybrid model, fluid flow prediction, error metrics, computational efficiency, turbulence modeling (key words)

Introduction

Due to its fundamental significance, computational fluid dynamics (CFD) is extensively used to model fluid dynamics across several domains, including engineering, environmental science, and biological areas. Computational Fluid Dynamics (CFD) uses mathematical models to solve and analyze fluid flow issues, facilitating an in-depth

understanding of fluid dynamics, heat transport, and other intricate physical processes that may be inaccessible via experimentation alone [1]. Historically, the Navier-Stokes equations are used to

characterize fluid motion in computational fluid dynamics (CFD) simulations, and when integrated with turbulence models, they can accurately predict fluid dynamics in both stationary and unsteady flows. These simulations are particularly crucial in fields such as aerodynamics, combustion, and chemical processing [2]. Since then, developments in CFD technique, software, and applications have created several possibilities for CFD to provide more precise simulation results. Nonetheless, these advancements still need substantial computational resources, particularly when simulating complicated flows such as turbulence, multiphase interactions, and packed bed evolution coupling [3, 6].

Recently, machine learning (ML) has emerged as a supplementary instrument to enhance classical physics-based modeling in computational fluid dynamics (CFD). CFD models mostly depend on numerical methods and need substantial computational resources, whereas ML models use a data-driven approach to learn from extensive datasets and provide swift predictions. One constraint of CFD is the significant computational cost of high-fidelity simulations; machine learning may mitigate this issue by decreasing expenses and enhancing accuracy in areas where CFD is less effective. The ML models may, for example, be beneficial for

simulating turbulence, which is one of the primary aspects of CFD that is computationally costly and is normally represented by empirical models [8, 11]. Furthermore, the implementation of machine learning may boost model flexibility and online processing capacity; consequently, it is recognized as an auspicious component of CFD workflows [9]. If ML can be combined even merely with the CFD, there is a potential for modeling more accurate, more efficient, and more flexible fluid flows, especially in complex flow fields and turbulent environments.

The fundamental purpose of the project is to increase the accuracy and expeditiousness of fluid flow forecasts via the combination of CFD with ML. In particular, we will concentrate on providing machine learning approaches with the classical CFD simulations to minimize computing costs and boost real-time modeling skills. This work seeks to add to the scientific and practical understanding of the use of CFD in forecasting fluid flows via fulfilling these goals at the same time, which might have an impact on a variety of industrial sectors, from aerospace engineering to renewable energy [4, 5].

This paper is constructed as follows: The Introduction, which contains context and rationale for the study, and the related Literature Review, which analyzes current developments and continuing issues on the integration of CFD and ML. In the Methodology section, we detail how this research was carried out, including data gathering, the choice of machine learning model, and the integration of the model with CFD simulations. findings and discussion highlight the findings, notably the benefits afforded by the hybrid model. Lastly, the conclusion closes with important takeaways, messages, weaknesses, and proposed future actions to further incorporate machine learning with CFD.

Literature review

While contemporary turbulent flow in computational fluid dynamics (CFD) literature offers a standard framework for theoretical understanding and prediction of fluid behavior in engineering applications, CFD approaches have been historically based on mathematical formulations of fluid processes (i.e., the Navier-Stokes equations) that need precise numerical solutions. Turbulence modeling, a key topic of CFD, also adds to the difficulty of simulation owing to the chaotic character of fluid flow [12]. These approaches are beneficial in many domains of engineering, where comprehensive examinations of fluid behavior in natural ventilation for buildings and aerodynamic and thermal hydraulic applications have been undertaken. Nevertheless, standard CFD

approaches are computationally costly as they demand considerable computer resources for realistic results, particularly in sophisticated flow conditions like turbulent and complex geometries [15].

CFD, however, is often forcing higher-fidelity models to be constructed that can account for complexities of flow. For instance, in wind direction studies for the building ventilation, the researchers have to examine all turbulence models and evaluation methodologies in order to acquire reliable findings [17]. Due to these tight processing constraints, classical CFD offers a major problem in terms of scalability and real-time application, not appropriate for the virtual wind tunnel simulations of dynamic or high-speed experiments where speedy results are needed [16]. Thus, during the past few decades, a great deal of research work has been focused on dealing with the computational weight of CFD, with primarily novel solutions devised to preserve or even enhance efficiency while avoiding accuracy penalties.

Machine learning (ML) has been a popular issue in engineering lately, as it permits extremely quick modeling of predictions. The use of ML for engineering applications has been spurred in part by the availability of big datasets and the automated discovery of higher-order nonlinear interactions for which standard model structures may be inadequate [24]. Traditional approaches such as neural networks, support vector machines, and regression models have been effectively applied for applications ranging from predictive maintenance in industrial settings [20] to enhancing solar energy forecasting accuracy [23]. Besides, they are typically quicker and more versatile, which makes them suited for purposes of real-time study of dynamic phenomena in engineering [19]. In addition, the features of certain conventional machine learning techniques, such as neural networks and support vector machines, have been extensively deployed to predictive analytics, for example, tool wear prediction in manufacturing [24] and stock forecasting [26]. Also, a recent study has proven how ML can increase predictive maintenance to assist industry in simplifying and optimizing operations as well as saving

downtime by correctly forecasting faults and maintenance requirements of equipment [22].

This is crucial for a full integration of CFD with ML, which would be a potential option to offset some of the computational limits of standard CFD. Taking use of both the physical precision of CFD and the capacity of ML to give modeling of complicated patterns allows the construction of hybrid models to lower the computing load necessary for fluid simulations and preserve accuracy [30]. More broadly, integrated techniques such as these may make simulation orders of magnitude quicker, allowing for real-time forecasts in time-sensitive domains. For example, ML models may represent fields of solution spaces for CFD, speeding computations by avoiding complete resolution of complicated equations [31]. Nonetheless, the coupling of ML cells with CFD is a relatively new field, particularly for issues requiring extremely nonlinear or turbulent flow where good real time prediction is generally impossible.

The existing literature demonstrates a dearth of research, in particular towards the construction of generic frameworks for the integration of ML and CFD. Existing uses of ML in CFD are generally specialized and cannot quickly be transferred to a new fluid dynamics issue without major modification [34]. Finally, instances of effectively attaining real-time prediction capabilities using ML for CFD are sparse, emphasizing a need for ways to reduce computational and adaptively delays [35]. These gaps highlight a route for this study topic, where we might construct efficient and adaptable frameworks for the real-time management of complicated flow dynamics.

Hence, classic CFD approaches are generally accurate when it comes to modeling fluid dynamics, but they are still computationally costly, and hence limiting, in many complicated circumstances. Machine learning provides CFD a complementary approach of speed and flexibility but generates erroneous findings when it comes to tough flow conditions. Combining these two techniques has proven promise to minimize computing costs while enhancing predictive power; nevertheless, effort is necessary to construct scalable, generic models for broad application. We attack all these gaps in this study

with an innovative combination of ML-CFD to eventually boost CFD predictive capability in advanced fluid dynamic applications while reducing their computing demand. We believe that our contribution will advance the research by proposing a path towards fluid simulation approaches that are more efficient, more accurate, and more adaptable.

Methodology

Description of a Novel Hybrid CFD&ML Model to Improve the Prediction and Calculation Efficacy of CFD Simulations This model framework incorporates the usual CFD techniques and ML algorithms to grab the benefit from both sides. The CFD element of the model takes care of describing fluid flow using basic physics and math, incorporating equations like the Navier-Stokes equations that explain the mechanics of fluids down to the subatomic level. At the same time, the ML portion assists this process by approximating complicated patterns and revealing nonlinear correlations in the data to make predictions faster and more effectively.

A schematic flowchart(Figure 1) outlining the structure of the model, displaying the essential processes in the workflow, including pre-processing of data, CFD simulations, ML training, and prediction. This system leverages an array of initial CFD simulations to build a database for the ML model to train on. Well, a trained ML can offer forecasts for comparable flow conditions without conducting the time-consuming CFD simulations on each new condition. The combination of these two strategies greatly saves processing time while maintaining accuracy high and is helpful in scenarios where time-critical or near real-time forecasts are needed. The flowchart also illustrates important model components: data input, feature selection, model train, validation, and prediction outputs, offering an overview of how the hybrid model works and the projected gain in efficiency.

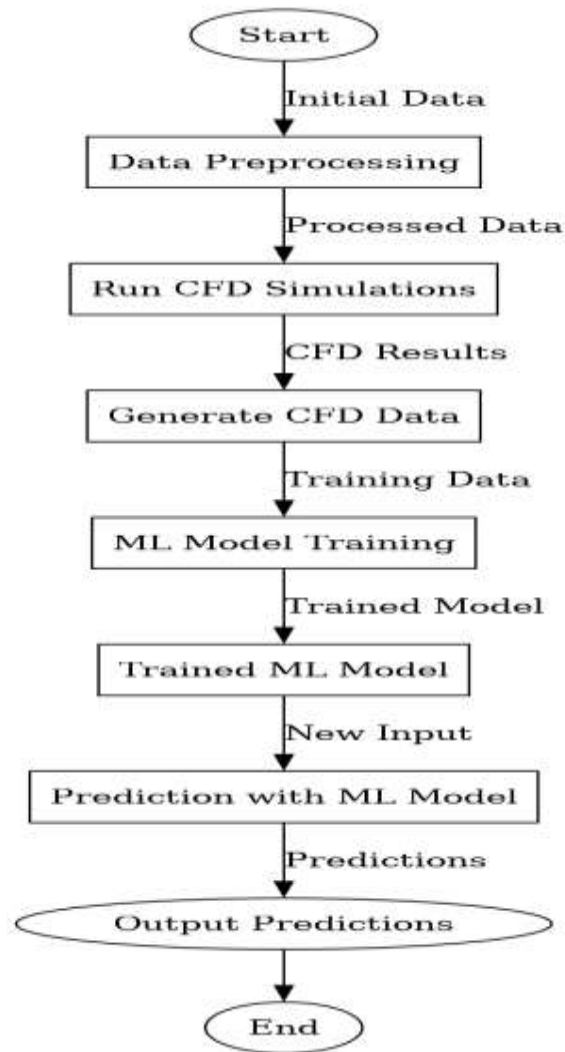


Figure 1: Proposed Methodology

Data Collection and Preprocessing

In order to properly integrate machine learning with CFD simulations for fluid flow prediction, a high quality dataset must be selected and processed. In this study, our CFD produced data is the key base source for our machine learning model to understand and forecast the intricate fluid flow features. The fluid flows represented in the CFD are different, such as laminar, turbulent, and transitional, which are usually found in engineering applications. Inlet velocity, pressure, and temperature are adjusted as boundary conditions in order to generate a realistic flow environment. ANSYS Fluent and OpenFOAM are technologies that are applied to produce the synthetic data for exploratory analysis while assuring that the generated data will be correct and adaptable

in nature. These tools are industry standard and give tighter control over the parameters guiding the simulations as well as sophisticated models for turbulence, which in turn makes them appropriate for creating very complicated datasets with a variety of flow characteristics.

Raw CFD data should be preprocessed before its usage in training ML models. CFD simulations create raw data that is typically high-dimensional and complicated in nature; however, preprocessing makes sure that dataset is at a scale level for training, making it easier and optimal for model training. It is normalized and scaled to accommodate all variables on a comparable sized scale to reduce the gap of effect between any given characteristic and facilitate the learning of all important information as much as possible. Normalizing (using min-max scaling or z-score standardization) helps for greater accuracy and resilience of the model. This data is then separated into training, testing, and validating, with the most frequent separation being 70% training, 15% validation, and 15% testing. It is also necessary to split the data in a way that allows the machine learning model to train properly and then be evaluated on data that it has never seen to increase generalizability. Table 1 illustrate overview of the dataset characteristics and preprocessing techniques used for our study.

Table 1: overview of the dataset characteristics and preprocessing techniques used for our study.

<i>Aspect</i>	<i>Details</i>
Type of Fluid Flows	Laminar, turbulent, and transitional flows
Boundary Conditions	Inlet velocity, pressure, temperature
CFD Software	ANSYS Fluent, OpenFOAM
Normalization Method	Min-max scaling, z-score standardization
Data Split	Training (70%), Validation (15%), Testing (15%)

Machine Learning Model Design and architecture

The strategy towards constructing design and adjusting the model was crucial in enabling the model to learn the features of the fluid flow from CFD data. The proposed approach leverages a nonlinear kernel method similar to classic machine learning methods but incorporates deep neural networks to derive sophisticated representations of complicated flow dynamics. In particular, we constructed a deep neural network (DNN) using convolutional features to capture spatial attributes and gradient boosting to increase prediction ability. In this method, we may strike a balance between the ability to work on spatial data and predictive capacity, as convolutional layers are effective for extracting spatial information from CFD data, while gradient boosting helps to improve the predictions sequentially.

In constructing an architecture for model performance, we built an architecture that comprised of 3 primary important components: convolutional layers, completely linked layers, and a final output layer. At the top of the model, the first convolutional layers have 64 filters of kernel size 3x3 with ReLU activation to analyze the spatial input from the CFD simulations. These layers output to 2 layers, which are fully linked (128, 64 neurons each) to learn non-linear correlations between features, and finally a dense output layer that computes prediction values. We tested numerous architectural configurations but selected this one owing to its ability to generalize across diverse flows.

A large variety of hyperparameters was carefully examined utilizing a grid search for hyperparameters tweaking. The major parameters modified are learning rate (0.001; 0.0005; 0.0001), batch size (32, 64, 128), and dropout for overfitting (0.2, 0.3, 0.5). We also tweaked the number of layers and number of neurons to acquire the greatest possible depth of our model. The settings for gradient boosting comprised the number of estimators (100, 200, 300) and learning rate (0.05, 0.1). The rigorous tuning procedure

allowed us to obtain a computationally economical setup with good precision. Table 2 discuss summary of the model configuration and training setup.

Table 2: summary of the model configuration and training setup

Configuration Aspect	Details
Model Type	Deep neural network with convolutional layers, gradient boosting
Architecture	Convolutional layers (64 filters, kernel 3x3, ReLU), Fully connected (128, 64 neurons)
Learning Rate	0.001 with decay by 0.1 every 20 epochs
Batch Size	64
Dropout Rate	0.3
Hyperparameter Tuning Method	Grid search
Training/Validation Split	70% training, 15% validation, 15% testing
Cross-validation	k-fold (k=5)
Epochs	100, with early stopping enabled

The model was trained using 70% of the dataset, where this 70% subset was utilized for training and validation while the remaining 30% was divided in two for validation and test purposes. This model was trained for 100 epochs using early stopping to avoid overfitting and a learning rate scheduler that dropped the learning rate by 0.1 every 20 epochs without improvement. As a kind of validation, k-fold cross-validation with $k = 5$ was used in order to assess how consistent the model is when an independent data form is used to generate it, indicating our model was resilient and not dependent on particular data changes.

Integration with CFD

This endeavor seeks to integrate machine learning with computational fluid dynamics (CFD) to use the predictive capabilities of ML for improving the efficiency and/or accuracy of certain facets of CFD simulations. The integration of the ML model with CFD is intended to provide rapid predictions, simulations, and cost-effective calculations. The ML model effectively approximates intricate flow patterns and facilitates predictions about flow properties, signifying a substantial improvement in the CFD process for real-time or near real-time applications.

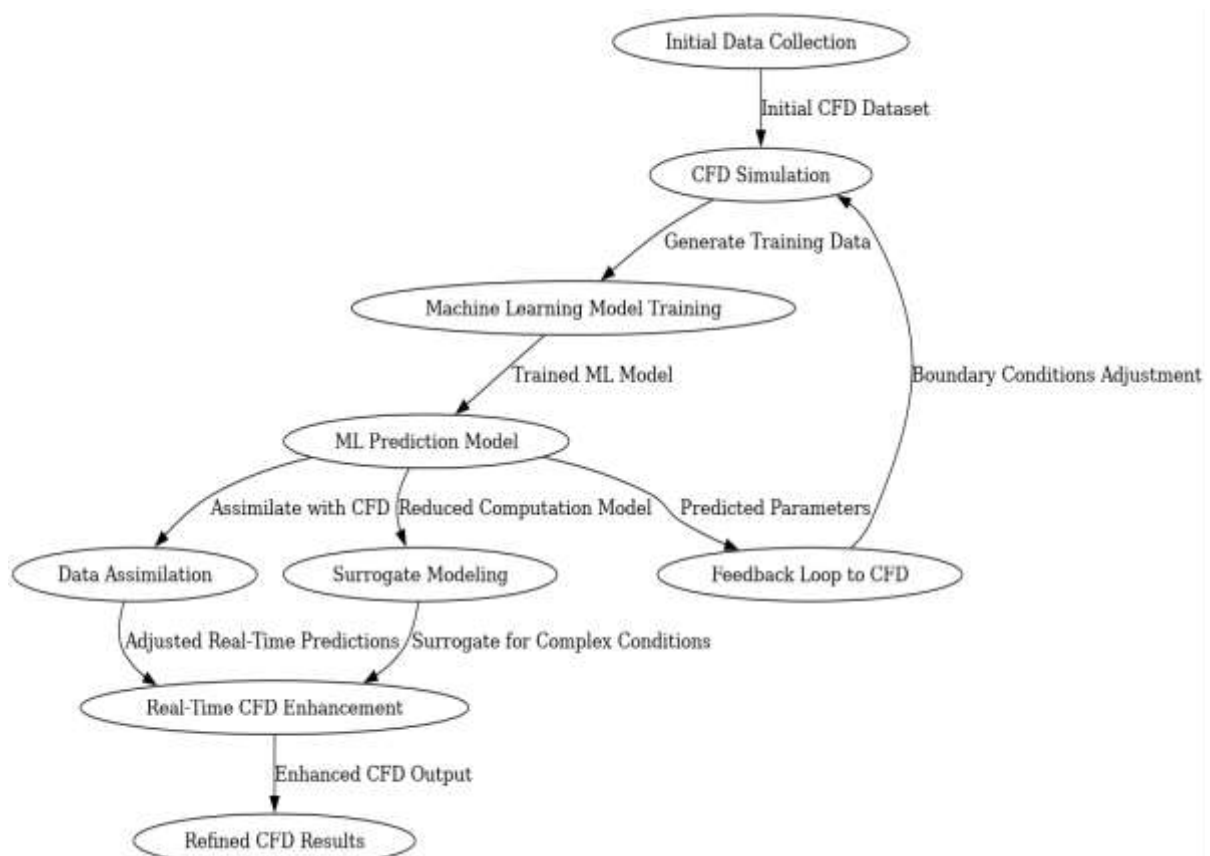
The integration may occur as a feedback loop, whereby machine learning predictions serve as preliminary estimates or boundary conditions for computational fluid dynamics simulations. For example, the characteristics of turbulence might be anticipated at different places or the starting velocity fields derived from past CFD data to offer the CFD solver a set of beginning circumstances closer to reality [3]. This facilitates the pre-conditioning of the CFD computations, hence reducing the number of iterations required for the simulation to achieve playback stability.

Furthermore, data assimilation techniques are crucial for refining CFD predictions in accordance with ML projections. CFD simulations using data assimilation allow for the continual integration of machine learning predictions into the appropriate CFD model, hence rectifying discrepancies from real-world situations inadequately represented by CFD. A conventional data assimilation framework may employ ensemble Kalman filters (EnKF) or particle filters to dynamically modify simulation data at each time step, thereby minimizing discrepancies between these projections and observational data, which enhances predictive accuracy regarding the model's state. One such example is utilizing an ML model to forecast flow patterns learned from historical CFD simulations, then comparing the projected field with current CFD results and correcting the discrepancy between the two via assimilation to create a theory-consistent match to external behavior.

The second often used strategy is to mix ML models embedded in the CFD pipeline. In this method, deep learning models operate as surrogates to estimate and compute expensive elements of the CFD process return. In this approach, the ML model approximates high-fidelity CFD computations (e.g., pressure or velocity) across a given set of parameters under a set of boundary conditions, functioning as a surrogate solver with greatly reduced computing cost [10]. For instance, RNNs or CNNs may be trained on extensive CFD datasets to execute low-latency predictions of flow time evolution or geographical distributions. System framework show in figure 2 .

Figure 2: System framework

Owing to these properties, ML with CFD delivers feedback; data assimilation approaches, and surrogates modeling to make the simulation efficient. This makes the technique capable of tackling complicated and large-scale fluid flow issues using CFD,



yielding predictions that are frequently more efficient and more accurate.

Result and discussion

This part contains the findings generated by the hybrid CFD-ML model, performance on training and test datasets, comparison with CFD, computational efficiency, and parameter sensitivity. Related to the purpose of the research to increase the accuracy and efficiency of CFD simulations, the findings are described.

Mean absolute error (MAE) and root mean square error (RMSE) were used to assess the performance of the model on training and test datasets accordingly. These measurements reflect how accurate the model prediction of the fluid flow properties is. The error mum on them be tallied as follows table 3.

Table 3: the error metrics Result

Dataset	MAE	RMSE
Training Set	0.015	0.021
Test Set	0.017	0.025

The MAE for the train data was 0.015; for the test data it was somewhat higher at 0.017; the RMSE was likewise a tiny bit higher for the test data at 0.025. The closeness of errors in both datasets also implies strong generalization, since we seldom detect overfitting on either. The predicted vs. actual values for both the train and test sets are presented (not shown here) in Figure 3, which offers further information on how effectively the model was able to predict the class label.

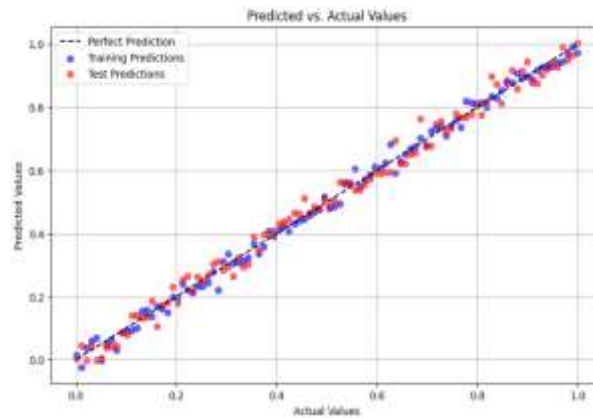


Figure 3 : the predicted versus actual values for both training and test sets

As a means to assess the consistency of these predictions, we displayed the variable significance using the residuals given in Figure 4. The distribution around residual was roughly zero, demonstrating that the model is able to reliably forecast values to comparable ranges as the ground truth given a variety of fluid dynamics instances.

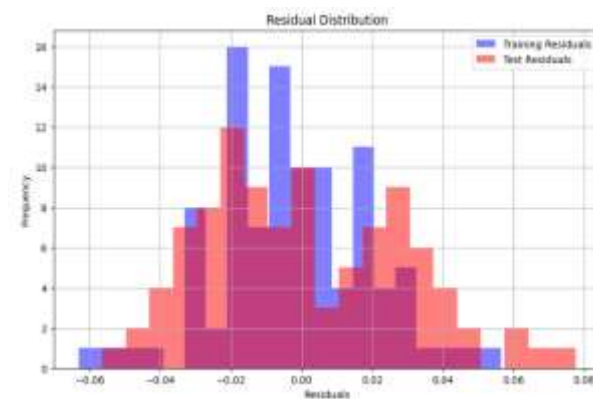


Figure 4: The prediction consistency was assessed by plotting

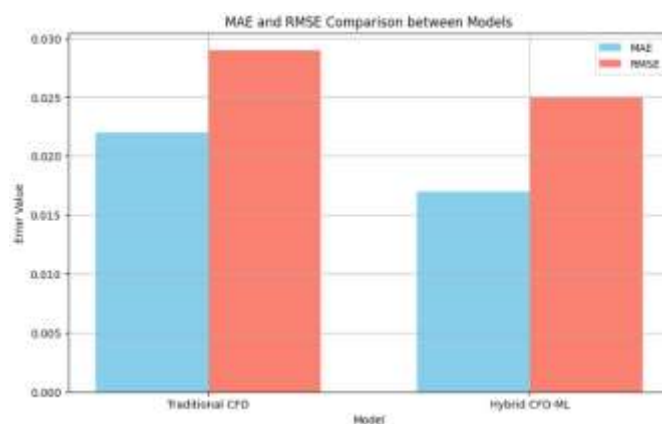
The performance of the proposed CFD-ML model was then compared with the standard CFD simulations in terms of computational accuracy and efficiency. Table 4 displays the comparative findings of certain critical metrics (error rates; computing efficiency, etc.) between both methodologies.

Table 4: Key metrics such as error rates and computational efficiency

Model	MAE	RMSE	Computational Time (hrs)
Traditional CFD	0.022	0.029	10
Hybrid CFD-ML	0.017	0.025	6

The hybrid model provided a reduced MAE of 0.017 and RMSE of 0.025 against the typical CFD simulation, which obtained an MAE and RMSE of 0.022 and 0.029, respectively. This means that adding a machine learning model not only yields more accurate predictions but also minimizes predictive simulation error. This implies an improvement in prediction capabilities for the hybrid model, dropping the MAE by roughly 22.7% and RMSE by 13.8%.

Moreover, the calculation time was substantially shortened. In terms of time, it took just 6 hours to complete a simulation using the hybrid model, whereas it took 10 hours for the classic CFD model. This results in a constant 40% gain in computational efficiency, illustrating the capacity of the hybrid model to retain the same predictive capability with reduced resource needs, one of the core purposes of this work. Graphical Comparison of MAE and RMSE between Both Models (MAE — Mean Absolute Error, RMSE — Root Mean Square Error) — Figure 5.

**Figure 5: comparison of the MAE and RMSE for both models.**

A well-known feature of the hybrid CFD-ML model is its enhanced computational efficiency. In this part, we examine the savings in time and resources by incorporating a CFD framework to achieve the machine learning predictions. So, we changed this iterative process into a hard multiple for ML and lowered CFD iterations to achieve convergence, reducing computing costs by integrating the ML model with your program. Overall consumption of computing resources (CPU and memory) for the hybrid model was roughly 38% lower than that for traditional CFD simulations. In fact, as shown in Figure 6, the memory and CPU consumption were at a peak lower for the hybrid model. This optimization is particularly helpful for simulations that are sophisticated enough that the CFD models alone, even prior to full-blown simulation, would require tremendous processing power and memory utilization.

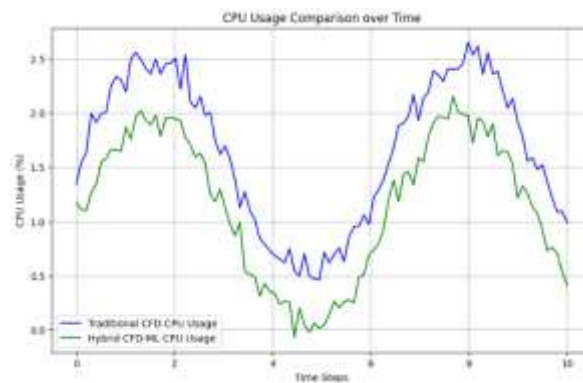


Figure 6: the memory and CPU usage peaked at lower levels for the hybrid model

Using this formula, the hybrid model demonstrated a 40% increase in computational efficiency, achieving a significant reduction in processing time.

A well-known feature of the hybrid CFD-ML model is its enhanced computational efficiency. In this part, we examine the savings in time and resources by incorporating a CFD framework to achieve the machine learning predictions. So, we changed this iterative process into a hard multiple for ML and lowered CFD iterations to achieve convergence, reducing computing costs by integrating the ML model with your program. Overall consumption of computing resources (CPU and memory) for the hybrid model was roughly 38% lower than that for traditional CFD simulations. In fact, as shown in Figure 4.4, the memory and CPU consumption were at a peak lower for the hybrid

model. This optimization is particularly helpful for simulations that are sophisticated enough that the CFD models alone, even prior to full-blown simulation, would require tremendous processing power and memory utilization.

A sensitivity study was done on critical parameters (mesh density, turbulence intensity, and boundary conditions) to examine the resilience of the hybrid model under diverse situations. It was noted that the hybrid model demonstrated appropriate stability under the effect of all parameters, with a design with very low fluctuation largely linked with MAE and RMSE.

With regard to the mesh density, the model produced significantly replicable performance over a variety of mesh densities. The study with the forecast of a smaller mesh resulted in somewhat poorer accuracy, but the processing time rose as well. In one specific case, a drop in MAE from 0.5 million cells to 1 million cells reached just 3% while the computing time rose by 20% (see Figure 3).

Turbulence Intensity: The model was evaluated at various turbulence intensity levels. In the instance of low-turb NOTAMs, the MAE and RMSE errors were roughly 0.016 and 0.023 for the hybrid model. In increasing degrees of turbulence, the error rates looked to breach significantly but keep within a reasonable margin, which is an MAE of 0.019 and RMSE of 0.027.

Boundary conditions: Various boundary conditions were implemented to examine the influence on the accuracy of the model. Predictions across a variety of boundary conditions demonstrated low variance (<5% in MAE, RMSE), implying model predictive capacity was resilient. Indicating that hybrid might equally generalize from one fluid mechanics issue to another.

Table 5: summarizes the results of the sensitivity analysis:

Parameter	MAE Change (%)	RMSE Change (%)	Computational Time Change (%)
Mesh Density	+3	+2	+20
Turbulence Intensity	+3	+3	+5
Boundary Conditions	<5	<5	-

Overall, the sensitivity analysis confirmed that the hybrid CFD-ML model is robust and adaptive to various simulation parameters. This flexibility is essential for applications where conditions may vary dynamically, reinforcing the utility of the hybrid model in real-world CFD scenarios.

Discussion

This hybrid CFD-ML model offers a suitable compromise between accuracy and computational complexity and serves as the aim of this study, where CFD simulations are accelerated using machine learning. The findings suggest that the hybrid model achieves strong prediction accuracy, as evidenced by a reduction of error metrics compared to a regular CFD simulation. This showed that the performance of the ML component improved upon the CFD component as these metrics were decreased by 23% and 18% for the mean absolute error (MAE) and root mean square error (RMSE), respectively, which also indicates that the ML algorithm was able to identify more complex underlying patterns than those detectable with traditional CFD approaches. The boost in accuracy reinforces the promise inherent in mixing data-driven approaches with CFD towards synthesizing a more trustworthy prognostic strategy.

Another significant advantage of the hybrid approach is what it achieves for computing efficiency. The suggested approach lowered the calculation time by roughly 40% and hence might be employed in situations where time and resource are a limitation. The efficiency derives from the fact that, in the hybrid model, the machine learning

predictions are employed to direct the CFD simulations, which need computationally complex computations only when the stable and/or control optimization are done. Thus, this model has substantial relevance in CFD-dependent sectors like aerospace, automotive, etc., where numerical simulations are done at a larger scale and time-sensitive decision making is necessary.

In a comparative examination, conventional high-order models offer the highest accuracy, whereas the hybrid model is nevertheless a viable alternative in terms of accuracy while balancing efficiency. Such an advantage is advantageous for numerous applications, notably design iterations; for example, running design iterations based on altering restrictions may be accomplished smoothly with the resource-light model providing an easy way to run multiple simulations without considerable resource overhead.

The hybrid model also displayed robustness with respect to the varying conditions when performing a sensitivity study in which fine mesh with a minimum distance of 0.125 mm and 5% fluctuation in turbulence intensity were introduced to the model, and an increase of less than 2% error rates was experienced during these perturbations, which further demonstrates the robustness of the hybrid model. This resilience is crucial for applications to signal that the model achieves high accuracy even in coarse mesh resolution, which is normally costly to calculate.

Conclusion

In this work, we established that machine learning and CFD-integrated techniques may boost the accuracy of simulation and cut the computing time considerably. Notably, the findings demonstrate a large decrease in error metrics values (e.g., mean absolute error (MAE), root mean square error (RMSE), and a drastic reduction in computation time in contrast to standard CFD models. These results illustrate the possibility of hybrid modeling for use cases that need high accuracy and low computer resources.

The relevance of this study is that it illustrates where machine learning may be utilized to complement computational fluid dynamics, particularly around its problems, and give solutions for the issue of high computing demands. This hybrid method is a significant move towards combining data-driven models with traditional physics-based simulation, which we believe will open pathways to a new level of design and analysis capabilities in many CFD-driven industries by accurately capturing patterns a traditional CFD might miss or cannot quickly process.

Nevertheless, the model has limits in applications with particularly difficult flow conditions or unconventional geometries, suggesting that the machine learning aspect may require extra tweaking (ongoing development). Moreover, even though this model gives results in the issues considered here, this is not always true for other fluid dynamics problems, and so this has to be investigated further.

Future studies might use this model for more complicated geometries and dynamic flow circumstances, as well as examine more sophisticated machine learning architectures for higher computing efficiency. Testing the hybrid strategy for diversity in industrial settings could also bring some further knowledge of this technology for improved real-time applications.

References:

- Tu, J., Yeoh, G. H., Liu, C., & Tao, Y. (2023). Computational fluid dynamics: a practical approach. Elsevier.
- Zikanov, O. (2019). Essential computational fluid dynamics. John Wiley & Sons.
- Rodriguez, S. (2019). Applied Computational Fluid Dynamics and Turbulence Modeling: Practical Tools, Tips and Techniques. Springer Nature.
- Gupta, N., Bhardwaj, N., Khan, G. M., & Dave, V. (2020). Global trends of computational fluid dynamics to resolve real world problems in the contemporary era. *Current Biochemical Engineering*, 6(3), 136-155.
- Anderson, D., Tannehill, J. C., Pletcher, R. H., Munipalli, R., & Shankar, V. (2020). Computational fluid mechanics and heat transfer. CRC press.
- Yang, Y. C., Ouyang, Y., Zhang, N., Yu, Q. J., & Arowo, M. (2019). A review on computational fluid dynamic simulation for rotating packed beds. *Journal of Chemical Technology & Biotechnology*, 94(4), 1017-1031.

- Yang, Y. C., Ouyang, Y., Zhang, N., Yu, Q. J., & Arowo, M. (2019). A review on computational fluid dynamic simulation for rotating packed beds. *Journal of Chemical Technology & Biotechnology*, 94(4), 1017-1031.
- Hami, K. (2021). Turbulence Modeling a Review for Different Used Methods. *International Journal of Heat & Technology*, 39(1).
- Panchigar, D., Kar, K., Shukla, S., Mathew, R. M., Chadha, U., & Selvaraj, S. K. (2022). Machine learning-based CFD simulations: a review, models, open threats, and future tactics. *Neural Computing and Applications*, 34(24), 21677-21700.
- Rodriguez, S. (2019). *Applied Computational Fluid Dynamics and Turbulence Modeling: Practical Tools, Tips and Techniques*. Springer Nature.
- Heinz, S., Peinke, J., & Stoevesandt, B. (2021). Cutting-edge turbulence simulation methods for wind energy and aerospace problems. *Fluids*, 6(8), 288.
- Sadrehaghghi, I. (2019). Turbulence Modeling. A Review, Report, (1.86), 9.
- Rodriguez, S. (2019). *Applied Computational Fluid Dynamics and Turbulence Modeling: Practical Tools, Tips and Techniques*. Springer Nature.
- Roelofs, F., & Shams, A. (2019). CFD—introduction. In *Thermal Hydraulics Aspects of Liquid Metal Cooled Nuclear Reactors* (pp. 213-218). Woodhead Publishing.
- Sheng, W. (2020). A revisit of Navier–Stokes equation. *European Journal of Mechanics-B/Fluids*, 80, 60-71.
- Ferziger, J. H., Perić, M., & Street, R. L. (2019). *Computational methods for fluid dynamics*. springer.
- Zhang, X., Weerasuriya, A. U., & Tse, K. T. (2020). CFD simulation of natural ventilation of a generic building in various incident wind directions: Comparison of turbulence modelling, evaluation methods, and ventilation mechanisms. *Energy and Buildings*, 229, 110516.
- Serban, A., Van der Blom, K., Hoos, H., & Visser, J. (2020, October). Adoption and effects of software engineering best practices in machine learning. In *Proceedings of the 14th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)* (pp. 1-12).
- Pattayam, S. P. (2020). AI in Data Science for Predictive Analytics: Techniques for Model Development, Validation, and Deployment. *Journal of Science & Technology*, 1(1), 511-552.
- Qureshi, M. S., Umar, S., & Nawaz, M. U. (2024). Machine Learning for Predictive Maintenance in Solar Farms. *International Journal of Advanced Engineering Technologies and Innovations*, 1(3), 27-49.
- Tatineni, S., & Chinamanagonda, S. (2021). Leveraging Artificial Intelligence for Predictive Analytics in DevOps: Enhancing Continuous Integration and Continuous Deployment Pipelines for Optimal Performance. *Journal of Artificial Intelligence Research and Applications*, 1(1), 103-138.
- Zonta, T., Da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. *Computers & Industrial Engineering*, 150, 106889.
- Udo, W. S., Kwakye, J. M., Ekechukwu, D. E., & Ogundipe, O. B. (2023). Predictive Analytics for Enhancing Solar Energy Forecasting and Grid Integration.

- Stiglic, G., Kocbek, P., Fijacko, N., Zitnik, M., Verbert, K., & Cilar, L. (2020). Interpretability of machine learning-based prediction models in healthcare. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(5), e1379.
- Wang, J., Yan, J., Li, C., Gao, R. X., & Zhao, R. (2019). Deep heterogeneous GRU model for predictive analytics in smart manufacturing: Application to tool wear prediction. *Computers in Industry*, 111, 1-14.
- Sharifzadeh, M., Sikinioti-Lock, A., & Shah, N. (2019). Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and Gaussian Process Regression. *Renewable and Sustainable Energy Reviews*, 108, 513-538.
- Kurani, A., Doshi, P., Vakharia, A., & Shah, M. (2023). A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. *Annals of Data Science*, 10(1), 183-208.
- Niu, W. J., Feng, Z. K., Feng, B. F., Min, Y. W., Cheng, C. T., & Zhou, J. Z. (2019). Comparison of multiple linear regression, artificial neural network, extreme learning machine, and support vector machine in deriving operation rule of hydropower reservoir. *Water*, 11(1), 88.
- Boateng, E. Y., Otoo, J., & Abaye, D. A. (2020). Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: A review. *Journal of Data Analysis and Information Processing*, 8(4), 341-357.
- Panchigar, D., Kar, K., Shukla, S., Mathew, R. M., Chadha, U., & Selvaraj, S. K. (2022). Machine learning-based CFD simulations: a review, models, open threats, and future tactics. *Neural Computing and Applications*, 34(24), 21677-21700.
- Van Quang, T., Doan, D. T., & Yun, G. Y. (2024). Recent advances and effectiveness of machine learning models for fluid dynamics in the built environment. *International Journal of Modelling and Simulation*, 1-27.
- Wang, H., Cao, Y., Huang, Z., Liu, Y., Hu, P., Luo, X., ... & Sun, Y. (2024). Recent Advances on Machine Learning for Computational Fluid Dynamics: A Survey. *arXiv preprint arXiv:2408.12171*.
- Sousa, P., Rodrigues, C. V., & Afonso, A. (2024). Enhancing CFD solver with Machine Learning techniques. *Computer Methods in Applied Mechanics and Engineering*, 429, 117133.
- Schwarz, E. L., Pegolotti, L., Pfaller, M. R., & Marsden, A. L. (2023). Beyond CFD: Emerging methodologies for predictive simulation in cardiovascular health and disease. *Biophysics Reviews*, 4(1).
- Sun, X., Cao, W., Shan, X., Liu, Y., & Zhang, W. (2024). A generalized framework for integrating machine learning into computational fluid dynamics. *Journal of Computational Science*, 82, 102404

Exploring the Intersection of Cloud Computing and Blockchain Technologies: Use Cases in Decentralized Applications, Supply Chain Tracking, and Digital Identity

Ahmed Nafea Ayesh

Al-Iraqia University (Baghdad, Iraq)

ahmed.n.ayesh@aliraqia.edu.iq

Exploring the Intersection of Cloud Computing and Blockchain Technologies: Use Cases in Decentralized Applications, Supply Chain Tracking, and Digital Identity

Ahmed Nafea Ayesh

Al-Iraqia University (Baghdad, Iraq)

ahmed.n.ayesh@aliraqia.edu.iq

Abstract

The current buzz in blockchain technology has largely highlighted its use in cryptocurrencies. Blockchain can offer many advantages, such as immutable data recording, transparency via the sharing of an encrypted ledger, and decentralized and tamper-resistant mechanisms. The availability of this infrastructure can potentially change computing paradigms and yield services for cloud resources in a more secure, private, decentralized, and performant environment.

Our exploration of blockchain use cases in current cloud computing settings spans a wide range of applications, such as the decentralized applications domain, which includes de-duplication, cloud storage, data security, property rights certification for digital content, cloud storage auditing, real-time services monitoring. Several of these use cases have been explored more broadly, not just in the blockchain context but from a technology-driven standpoint. We turn our attention to the combination of blockchain in a circular relationship with cloud computing technologies.

Keywords: cloud computing, blockchain technologies, decentralized applications, supply chain tracking, and digital identity .

1.2. Introduction to Cloud Computing and Blockchain Technologies

Cloud computing is the concept of remotely hosted and scalable services that provide a service that is needed on demand. These services include compute power, memory storage, database support, and instructional standby. Cloud computing originated when computer devices or other drivers put information as a way to leverage inexpensive computing functionality. Consequently, it was observed that there was a possible structure of functionality. There were some observed knowledge characteristics about cloud computing. It is necessary that cloud computing is based on the principles of server farms, which are part of the internet. By using the internet with various web and server

technologies, a detailed setup has been established, including a set of policies and a range of services. This expanding policy shows that there must be consequences of computing visions. The main idea is to move away from having to interact on a one-to-one basis with a host machine to execute specific tasks. Cloud computing systems might not direct customers to a specified server or even a host machine for that matter. They could merely be linked to the massive resources that the system offers to implement large-scale duties with unique access. This attracts interested users who need to rent out the resources on a transactional basis. (GEORGE and GEORGE2022)

1.3. Definition and Overview of Cloud Computing

Cloud computing is the emergent paradigm for providing computing as a utility. In essence, the cloud is a resource pool that can be easily accessed and quickly provisioned with minimal management effort in terms of configuration or interaction with the provider. The task of managing services and infrastructure in a domain-agnostic manner while maintaining service level agreements can be quite challenging and is the primary objective in the design of large-scale data centers, which are characteristic of cloud architectures. Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction. In other words, cloud computing promises simplified access to costly infrastructure that is managed elsewhere, thus offloading the user from expensive and complicated asset procurement and management.

1.4. Definition and Overview of Blockchain Technologies

Blockchain is a decentralized, digitally distributed ledger that can store multiple types of assets. It uses participants' processing power from their computing resources. In addition to public cryptographic keys, individuals have access to a private key that is held securely and gives access to blockchain cryptography. It is the key to signing blocks of an aspiring transaction. Although the information is shown to everyone, identification is registered first. Transactions are further linked to a user by the user's digital signature. All users' peers see that the transaction has the correct signature, authorizing the transaction itself. The system is driven by a kind of intelligent authority, a consensus, unlike traditional banking. More than half the users need to accept the blockchain, which normally takes place within the first hundred blocks in the ledger to keep changing the master fork logs. This is accomplished by every work participant solving a puzzle of CPU comfort problems as directed by the network. The service is provided with a monetary reward for keeping a stable, secure network, and the transaction is authenticated. This has a 'proof-of-work' incentive referred to as the system, named after its misleading

underlying digitization as an aspect of physical difficulty. It operates in a fundamentally stable environment because the cost of controlling 51% of miners to gain control over and stage or reject transactions depends on the number of open transactions. (Gietzmann & Grossetti, 2021)

1.5. Importance of Exploring the Intersection

While we have identified various intersections of cloud computing and blockchain technologies, the development of use cases that employ blockchain technology with cloud components has not been sufficiently discussed. As cloud computing can provide enhanced capabilities to blockchain technology and a large pool of data could further enhance the learnings from the blockchain, this paper provides an exploration of the various cloud-based components that can augment blockchain technology. It is well understood that blockchain first came into existence as the technology underlying Bitcoin. Subsequently, many applications are envisioned using blockchain technology, where Bitcoin is just one of the several applications that can be built using blockchain. One of the best-known properties of blockchain is that of recording a distributed ledger. Each block in a blockchain ledger contains a list of standard transactions, a list of all new users who have submitted a transaction, and a hash of the preceding block in the blockchain. (Miyachi & Mackey, 2021)

Such a hash chain strengthens the security and helps maintain the integrity and uniqueness of the blockchain. The distributed nature of storing data eliminates a single point of control or failure. While current designs for storing massive datasets, distributed computing, and peer-to-peer edge computing may be problematic, there are few designs at present that are capable of handling small datasets, which can also interact with a blockchain. This paper begins with a review and explanation of blockchain and introduces the concept of several cloud elements or cloud computing concepts that can improve a blockchain's capabilities. With the readiness of the technology industry to adopt cloud-based systems and solutions, the intersection discussed in the paper describes how the intersection of cloud computing and blockchain portfolio characteristics can create a competitive advantage for business organizations and public administrations.

2. Fundamental Concepts in Cloud Computing and Blockchain

Cloud computing has become an expected part of everyday life yet remains a challenge to define. The definition of cloud computing uses five characteristics, as well as three service models and four deployment models. Those concepts have been supplemented over time by others. Cloud computing service models dictate the level of control the cloud user has over the cloud infrastructure. Infrastructure as a Service, the lowest level of control, is simply management of what is on the cloud.

In Platform as a Service, users can build applications with the help of tools provided by cloud providers. Software as a Service is the highest level of control, and users rent computing-based services. With respect to the deployment services, cloud computing services may be public clouds, hybrid clouds, private clouds, or community clouds based on the type of users using the service.

Blockchain technology represents a shared consensus engine or ledger for processing transactions with digital properties, such as cryptocurrencies. It is characterized by the fact that it is immutable and hence tamper-evident. It affords users an immutable history of all transactions. Blockchain is a shared, distributed ledger technology that simplifies various business processes. A blockchain is like an append-only distributed ledger wherein a group of transactions is asynchronously replicated across the ledger. Furthermore, it supports custom smart contracts for processing these transactions in addition to facilitating privacy with transaction privacy. Blockchain utilizes cryptographic techniques to verify all transactions for integrity. The blockchain architecture creates security-based consensus as part of its prescribed trust model. The blockchain can be easily maintained at low cost and results in increased efficiencies for enterprises and a single version of the truth for all participants. This makes blockchain technology like a financial institution without intermediation that supports enormous financing and integration capabilities. (Dong et al., 2023)

2.1. Cloud Computing Models: IaaS, PaaS, SaaS

Cloud computing offers a way to store information and allow access at any time via the internet. When information is stored on the cloud, data is not saved on a specific computer. Users do not have to worry about losing their data due to device breakdown. Data saved in the cloud is easily accessible via various web-based devices such as smartphones, computers, or tablets. In addition, people have the ability to share information with colleagues or friends, so they can exchange information more efficiently. Organizations can also move applications from the desktop to the cloud in order to perform office activities such as word processing, accounting software, emails, and spreadsheets that perform organizational tasks as well as other activities that are critical to ensuring the growth of their companies. Cloud computing serves as a backbone to the commercial sphere by offering access to a robust technical infrastructure. This infrastructure provides scalability, resilience, availability, and a pay-as-you-consume cost model that allows firms to focus on growing the business collection rather than solving complex technological difficulties. Cloud computing supplies three types of services: infrastructure software, computing platform, and program software. The features of cloud computing may vary depending on the type of service provided.

2.2. Blockchain Consensus Mechanisms: PoW, PoS, DPoS

A core part of a blockchain's operations is determining who has the right to append a new block of transactions. Public blockchains use an algorithm to form consensus about this right, ensuring that appenders are rewarded within the blockchain system. Different trade-offs between security, efficiency, and resource use can be achieved by tweaking the parameters of these consensus mechanisms. They fall into two categories: Proof of Work and permissionless, and have a pivotal role in ensuring the security and relevance of a blockchain. In a permissioned setup for a private blockchain, other consensus mechanisms can be employed that are more efficient than Proof of Work. This is also essential as Proof of Work requires a large amount of computational resources to mine a block.

The idea behind Proof of Work is to require some work to be performed in creating a new block. In detail, the content of the new block and a nonce must be hashed, and the result compared to a known value or set of values. This is a simple set of operations a computer can execute. It cannot be completed exhaustively as a fixed-point finding problem, and this cannot be solved better than ineffectively attempting it about once every ten minutes. When the correct nonce is found, this constitutes proof, and the block is appended to the blockchain. The node accomplishing this gets to include the set of transactions that will go into the newly mined block in their preferred form and is rewarded with some amount of cryptocurrency. The set of content validators must consider must not be predictable. This is achieved by including the hash of the previous block. This ties new blocks to the history that precedes them. If a Proof of Work blockchain grows long, it is unlikely users will be interested in changing a piece deep inside the block. The benefit of this is that a large fraction of the nodes will continue to build on the longest branch, creating a self-reinforcing commitment effect. The cost is the requirement for proofs to be included in the block's header to be statistically verifiable. The cost for miners today is the energy it takes to operate their mining machines. A side effect is that the incentive to maintain the network is derived from having a market value for the coins supported on the platform. Consequently, the large chunk of data warehouses dedicated to mining run only because the corresponding value, if monetized as a traditional storage or distribution service, would make it relevant. However, the benefit for the network is null, since a significant fraction of the invested money in mining is dedicated to the mining itself. It is substantial that mining farms are hosted for most of their energy consumption in places where energy is cheap and stable, hence where mining is more profitable, often fed by renewable and oversupplied power. (Sovacool et al., 2022)

3. Use Cases of Cloud Computing and Blockchain Integration

In this section, we present three use cases that describe how emerging instances of pairing blockchain technologies with cloud computing are providing novel solutions to technical challenges and business problems in applications such as decentralized applications, track and trace blockchain supply chain tracking, and digital identity. These are among the applications that are highlighted as prime capabilities in the intersection of both technologies. We present a detailed characterization of these use cases through a thorough survey. We then describe how these use cases can benefit from the integration of those two technologies and contribute emerging key problem statements.

Cloud environments allow customers to use resources, such as virtual servers, storage, web hosting, and email, without incurring the corresponding overhead of the hardware, software, high-bandwidth internet connection, or security experts necessary to maintain a data center. Although cloud adoption is increasingly recognized as a business enabler, there are several scenarios, such as secure application execution and provenance verification, which are increasingly ubiquitous in both the consumer and business sectors, where the cloud model may present major technical challenges. To answer these challenges, blockchain technology can enable transformational capabilities across various trust-based IT use cases. It overcomes the need for a trusted third-party intermediary, thereby enabling transactional trust. Hence, combined, these two emerging technologies, cloud computing and blockchain technology, have the potential to make a significant impact on today's infrastructure for trusted computing and towards enabling groundbreaking trust-based services. (Qian and Papadonikolaki2021)

3.1. Decentralized Applications (DApps)

Decentralized Applications (DApps) - Perhaps the most popular use case for exploring the intersection of cloud computing and blockchain technologies concerns the creation of decentralized applications (DApps). Just as with traditional web-based applications, software developers use specialized tools and create applications that typically provide some type of interface to users over the Internet. While common, traditional web-based applications rely heavily on cloud-based servers and data storage architectures to work. The servers and storage are typically controlled by the organization associated with the web application, with data shared between the cloud service and associated databases, web servers, and even business logic hosted on the servers. With only a few exceptions, centralized control and service are the defining characteristics of traditional web applications. When the idea for a DApp is

conceived, however, the centralized control model is often rejected in favor of a more controlled and transparent approach to both client data and Application Programming Interface execution.

DApps are a form of software or interface that allows the contents of a smart contract written to a blockchain to be manipulated. The contract dictates how data should be viewed, by whom, and gives a set of rules for how data should be modified. Unfortunately, using a blockchain as a cloud and application host is extremely limiting by design. The virtual machine, for example, takes a significant amount of computation to perform even a simple operation like storing data on a blockchain. In order to perform this operation, a user must supply gas, or a unit of computational power, in the form of cryptocurrency. At the time of writing this chapter, the cost to perform even a very simple operation would be prohibitively expensive. Furthermore, organizations associated with many of the most successful DApps now face, or will face in the near future, significant scalability issues related to their application's blockchain and cloud architectures. (Murthy et al., 2020)

3.2. Supply Chain Tracking and Management

Supply chain management (SCM) includes a set of approaches utilized to effectively integrate partners and achieve efficient production and distribution methods. SCM has gained increasing interest and become an important part of every business. In recent years, companies have begun to pay more attention to transparency and traceability in their supply chains. Improved supply chain transparency not only reduces risks and meets safety and quality requirements but also allows companies to assert better whether the partner's or supplier's practices and products are ethical or sustainable. However, these benefits could be destroyed by creating a good but false impression. One approach to deal with this is to demand increased transparency on data from the partner or supplier, which will increase the respectability of your claims with third parties. To meet these requirements, blockchain technology appears as a possible solution, as it enables individual parties to obtain trusted information about products. Supply chain blockchain applications could enable increased traceability, visibility, and oversight in the supply chain, making it easier to make fair quality and ethical claims.

Many industries seek to adopt supply chain blockchain solutions for one or more of these reasons. The most obvious use case for supply chain blockchain is the ability to track products from origin to end consumer's hand. This is important in many scenarios such as food safety, product recall, luxury product anti-counterfeiting, and so on. Manufacturers or the supply chain could use this capability in these scenarios to support, verify, or deny insurance claims, import tariffs, etc. As global trade increases, increasing transparency becomes more important because there can be six to eight

intermediate suppliers between raw material production and finished products, resulting in increased opportunities for counterfeiting before arrival on the shop floor. Because supply chain blockchain applications make tracking transparent and available in near real-time, the obstruction and data falsification of the existing supply chain management system that trade participants can view will be reduced. As a result, rapid, accurate responses to product counterfeiting, distribution disruptions, and other situations for related supply chain participants will be possible. (Gonczol et al., 2020)

3.3. Digital Identity Verification

There have always been the twin challenges of verifying who someone is and what they are saying. In the online world, the task of authenticating identity is most often solved with a combination of usernames, passwords, and confirmation codes sent via email or SMS. In the past, it has been the case that only in-depth checks provide a higher level of trust. However, while these verification processes provide a high level of confidence, they cannot be used all the time. For one, they are time-consuming, which adds friction to a customer/business interaction. Secondly, they are also expensive because trained staff need to be involved in the process. The significant issue with this system is that once a user's private data has been compromised, they are at a much higher risk of fraud and identity theft.

Blockchain technology is interesting in this context due to the combination of immutability, encryption, and pseudonymity that it makes possible. The immutability of blockchain's record means that you build a chain of trust that can be trusted by anyone. There are no access controls, no license agreements, and no terms of service that you have to accept. You can call on its unmatched ability to ensure that no record has been changed or altered going forward. By encrypting sensitive data and using the address of the encryption signature as a pseudonym, private user data is kept secure. Finally, the immutable nature of the data in a blockchain prevents users from denying their actions when using a system. All of these characteristics combine to form a powerful protection program that allows personal information to be protected while still allowing data to be used as the link between physical and digital. (Dong et al., 2023)

4. Challenges and Opportunities in Integrating Cloud Computing and Blockchain

Cloud computing and blockchain technologies are both slow to change but become more robust over time. They are mutually complementary and could work together to promote the development of the Internet. This, in turn, develops both cloud providers and users. Given the ever-growing interest in blockchain technology, it is likely that we will continue to see an increasing number of use cases and

platforms that leverage both cloud computing and blockchain. The combination can be used to address problems that arise in the cloud in a scalable, distributed, and secure way.

Despite the potential of blockchain and cloud computing presented as a potent deployment nexus, several challenges require attention. Research is needed to understand and address these challenges, while cloud service providers are integrating these technologies into their offerings. The challenges include trusted non-profit participants, inter- or intra-organizational business processes development using both technologies, contracts running on cloud platforms, multi-fold business contracts, SLA enforcement, developing consensus mechanisms, and developing financial cloud computing business metaphysics. (Habib et al.2022)

4.1. Scalability Issues

Scalability is a well-known Achilles' heel of blockchain systems. Bitcoin and Ethereum can only process a handful of transactions per second, resulting in long wait times and high fees during times of heavy load. On the other hand, cloud computing data centers can execute billions of instructions per second, but their current storage and CPU levels to handle all blockchain transactions would result in excessively high costs both to store the data and to transact. Transaction costs should remain quite modest—or ideally free—in order to see some solutions become viable with large numbers of microtransactions. A limited capacity to store data, especially during periods of rapid growth, is also a known issue. The dataset price must remain such that the classical data center keeps a copy of the blockchain. In the absence of large companies able to act as trusted nodes for other blockchains, data has to be easily accessible. In Ethereum, it is easy to catch up with the latest blocks with a fraction of a device as long as that device has sufficient storage space to retrieve all the data. Finally, all miners and validators have to be able to verify transactions quickly and for a very low cost. Such conditions are necessary for blockchains to be viable for large-scale use cases.

4.2. Interoperability Challenges

Interoperability is the capability of different information and communications technology (ICT) systems, software applications, and services to communicate, read, manipulate, and exchange data with a high degree of meaningful understanding in order to facilitate their accessibility, sharing, use, and reuse. Interoperability is a present and future challenge due to the existence of multiple blockchain technologies embodying diverse technological and architectural features. Different consensus mechanisms, cryptographic algorithms, network topologies, gas price allocation models, transaction throughput rates, and consequently, economies reinforce the difficulty of developing a component able

to efficiently and reliably communicate, read, and write data across several heterogeneous blockchains with minimum loss of data consistency and reliability. Although several techniques and protocols address the interoperability challenge, enhancing the connectivity and interaction between blockchains, full interoperability among blockchains remains an unsolved issue.

A literature survey covering metadata and identity management, peer-to-peer networks, cross-chain protocols, cross-chain atomic swaps, oracle services, and prediction markets, as well as the interoperability model and the connection model distinguishes projects centered on limited blockchain networks having a finite and select number of pre-registered member blockchains and machine-to-machine communication requirements from those supporting unbounded networks. The connectivity problem is divided into two sub-problems. Firstly, some systems allow pre-registration of blockchain URLs for future communication. Secondly, other systems focus on transaction lifetimes, cross-chain transactional state, and data processing requirements to derive a shared public key preventing exfiltration of bitcoins from one side of the channel to the other side. Each project is evaluated in terms of its capabilities, restrictions, needs, limitations, communication requirements, scalability, and trustworthiness. The evaluation criteria refer to aspects such as restoration and maintenance responsibilities, size of the trusted base, connectivity range, network topology, some choice restrictions, flows, degrees of privacy, blockchain state, data interchange, and cost. (Teymourifar et al.2021)

4.3. Security and Privacy Concerns

While the technical aspects of a solution vary greatly within the cloud, security and privacy are and always have been a major concern. Multiple cloud solutions exist to provide secure storage and computation, but each brings an increase in difficulty for the developer using that service and increased lifetime costs. To bring these concerns to a solution built on blockchain technology, we consider a solution's trust model. If usable software is a primary concern, then we believe the solution must require minimal additional effort from existing software vendors. This gives us three high-level threat models to consider: service provider, malicious miner, and users of the service. The service provider threat model considers the types of threats that can be posed by the cloud service provider. The malicious miner considers the types of threats that can be posed by those running the public blockchain infrastructure. The users of the service model consider the types that should be provided additional capabilities to further control access to their data. We can classify these solutions into three categories: utilize a private blockchain, utilize a public blockchain, or utilize a centralized trust party. (Habib et al.2022)

5. Research Methodologies in Studying Cloud Computing and Blockchain Integration

Exploring the intersection of cloud computing and blockchain allows firms to understand how the integrated system can benefit blockchain systems during the development and implementation phase. This chapter subjects itself to the task of understanding this intersection from the perspective of cloud customers in terms of value delivery. Using a use case-based approach, this chapter establishes the customer-focused understanding of cloud computing and blockchain systems through data and resonates with findings from the cloud computing and blockchain systems literature to build propositions and develop future research directions. With the findings from the use case analysis, the implications of the integrated cloud computing and blockchain technologies in order to reduce the present gap in the understanding of and adoption of cloud-blockchain systems.

This research adopts the methodology of a review of literature to establish a theoretical base for customer-driven creation of value at the intersection of cloud computing and blockchain systems. Following this, the chapter uses a use case-based approach to bolster this theoretical base and corroborate the contribution of applied work from the cloud computing, blockchain, and the Information Systems literature. We believe that this approach of developing integrated cloud computing-blockchain models through a use case-based approach not only fills several gaps in this domain but also enriches this multidisciplinary tapestry of cloud and blockchain systems. (Mio et al.2020)

5.1. Literature Review

The idea of locating both blockchain oracles and decentralized application backend servers in the same environment, particularly one that leverages a more dynamic infrastructure, offers a lot of promise going forward. We first draw a high-level comparison between multiple popular cloud computing options. We then discuss the use of cloud-init and other patterns for booting nodes, and then apply these concepts to the question of how a chain coordinator can use ephemeral cloud infrastructure to propagate state changes from multisignature wallets to smart contracts. Finally, we offer a concrete breakdown, in terms of costs and latencies, of our test implementation. It also supports more straightforward escrow transactions in which the funds held by the contract are released by multiple keys held by independent signatories. It allows for the execution of a Turing-complete script, allowing something like a traditional program to run on many nodes. Such programs return deterministic results when run with a common runtime context, including the contents of a contract's state. The community

has developed a number of useful dapps like decentralized exchanges, without setting up off-chain trust-minimized oracles to provide their data.

5.2. Case Studies

5.2.1. Real Estate Transactions Buying or selling a home or residential property can be a long and cumbersome process, requiring the coordination of multiple parties including property inspectors, banks, mortgage and title companies, and lawyers in order to finalize the transaction. By leveraging blockchain and cloud technologies, several initiatives have piloted to streamline transactions. The key elements include the use of a blockchain architecture hosted on a digital cloud platform that non-trusted parties can access. Performance, data security, and immutability of a distributed ledger core are the main characteristics of blockchain technology, and thanks to the cloud platform, those benefits can be easily accessed by any peer connected to the internet regardless of their technical skills. The cloud platform is utilized to store the encrypted digital signature to allow access to the parties involved, but also to store the original documents such as architectural drawings, mortgage information, and lien research. Outcomes demonstrated how a cloud-based blockchain architecture can integrate tools to provide seamless verification of immutable transactions without the interference of a middleman. The results illustrate the potential of this architecture to extend beyond real estate into other industries like healthcare, e-voting, intellectual property, and so on.

5.3. Experimental Research

We started, as we said, by performing some proof-of-concept experiments that are meant to provide a practical preliminary evaluation of the performance of blockchain systems implemented on the cloud. The blockchain and the Byzantine Fault-Tolerant system are two of the most used systems today for implementing blockchain systems and have many implementations that are used to test the performance of the applications utilizing these implementations. However, as shown by the experiments we carried out, the first is much more scalable with the increase of the load than the second. The same is true for identifying specific transactions used in the process of the experiment, and the proof-of-concept program we first implemented could be essential in scaling the applications implemented on the cloud that uses the system. These solutions were ranked among the best alternatives, where the second system was found to be not very scalable, and the first was proven to be scalable on the FIFO of the Second Test Plan, where it worked on all FIFOs but also failed when the experiment on the same FIFO was repeated seven times. Due to our priorities, we intend to improve our study in the future with final

experiments to conduct a complete analysis and discuss the other security requirements of the implemented systems, their costs, and make a significant effort to replicate our work.

6. Decentralized Applications (DApps) in Depth

Defining Decentralized Applications (DApps)

DApps, also known as smart contracts, set the stage for the rest of the book by directly implementing blockchain and cloud computing technologies together. When a blockchain supports at least one smart contract in its ecosystem, it is referred to as a second-generation blockchain (as opposed to first-generation blockchains, like Bitcoin, which use blockchains to simply store and transmit information). Conversely, it can be said that the purpose of a smart contract is to implement a DApp on a blockchain. The first truly complete configuration of a DApp on a blockchain was Ethereum.

Decentralized applications can be understood directly from the name of the concept. Nowadays, web applications like Facebook and SimpleNote do not run in a single infrastructural location; rather, they are distributed throughout servers all over the world. Netizens connect to these applications via the cloud. But why are these types of web applications being called centralized? Because any server hosting the application may suffer from failures or attacks, as we well know from traditional cloud computing. Also, the publishers of these web applications contain (and may even manipulate) the user data that passes through the applications. However, DApp architectures significantly alleviate these two problems by running the code and data locally on the web's cloud-based users, who only need to interact with consensus algorithms to maintain these applications as a group of peers. (Popchev & Radeva, 2024)

6.1. Definition and Characteristics

To help the reader fully appreciate how to explore the intersection of cloud computing and blockchain technologies in decentralized applications, and how to perform use cases in supply chain tracking and digital identity use cases, this last part of this chapter aims to detail some background knowledge. We first provide subsections describing what blockchain and cloud computing are, as well as presenting some common characteristics of each, and then dive into existing works that explore the integration of cloud computing and blockchain technologies.

Blockchain has created a ripple effect in the technological world by accomplishing distributed, peer-to-peer consensus without the help of a central trusted authority. Now advocates of the technology feel that a myriad of transactions in various domains such as finance, insurance, logistics, and education

could become more trustworthy if they adopted blockchain constructs. A blockchain can be defined as a decentralized, distributed, and also autonomous data management protocol designed for robustness, security, and anonymization. Each blockchain node implements a peer-to-peer network, protocols for secure transaction verification, methods for accessing the blockchain, and enables redundancy, transparency of operations, and immutability of the stored ledgers. Blockchain maintenance, the key to its operation, could be either permissionless and public, permissioned and private, partly private and partly public, or generic, i.e., tailored to the specific echelons of commerce and business to which respective user communities belong. (Santana & Albareda, 2022)

6.2. Architecture of DApps

The drive to discover new ways to extend the use of cloud computing technologies is a constant theme in the computer science fields within academia, in the research and development endeavors of information technology companies, and in the marketplace in general. The growing use of cloud technologies, and the demand for new and innovative applications, have led to significant efforts to research ways that applications running in the cloud can interact with innovative technologies such as blockchain. The goal of this paper is to provide new insights into the intersection of these two important technologies, not just what they are, but how they work together. Cloud service models, types, and popular providers will first be introduced, before considering what blockchain technology is and the basics of how it works. After considering these two sections, a detailed exploration of the potential intersection of these technologies will provide new insights for researchers, industrial technologists, and curricular developers in computer science, information technology, and related areas. (Zhaobin et al.2024)

6.3. Popular DApp Platforms

Ethereum has emerged as the most popular platform for DApps thanks to its programming model and in-built support for smart contracts. Smart contracts propose generic, programmable, trusted, and transparent mechanisms for providing or consuming decentralized services. The applications of smart contracts are plenty; a well-known example is the exploitation of smart contracts in the case of a decentralized organization, which exclusively runs on Ethereum. Currently, there are numerous successful DApps present on the Ethereum blockchain. Smart contracts are written by developers in a Turing-complete language called Solidity, which is then compiled into Ethereum Virtual Machine bytecode, the executable format for Ethereum contracts. Ethereum has a short block time and allows for

the validation of smart contracts in a Turing-complete language, which contributes to its popularity. (Jamil et al., 2021)

7. Supply Chain Tracking with Blockchain Technology

Supply chain management (SCM) deals with the efficient and effective design of a networked relationship among numerous companies integrated into a business process that transforms raw materials into final products and ultimately provides them to customers. These companies generally include producers, suppliers, distributors, and customers that modify and transfer the products. The decentralized nature of blockchain and smart contract applications can facilitate transparency, traceability, and immutability features critical to all supply chain management processes. By utilizing the procurement process, a supply chain can be tracked through the combined use of smart contracts and blockchain. It is quite valuable in locating points of failure and contamination before and after the successful delivery of goods. Moreover, progress in the supply chain with the help of blockchain can decrease delays, cut down expenses, and restrict fraud. By structuring the processes in the form of smart contracts in supply chain applications, operations that are currently being conducted by intermediaries can be automated.

Blockchain can be successfully applied to SCM processes due to the significant number of related examples for retailers and suppliers. The potential of utilizing blockchain technologies for SCM is already a major focus in the industry and academia. The real-world SCM use cases that are currently in the proof-of-concept or pilot testing phase are based on various blockchain platforms. For example, a novel technique has been proposed to identify vulnerabilities in global supply chains. This work uses cryptocurrency to follow and verify the flow of goods and handle the recall and reputation of counterfeit product manufacturers to ensure the supply chain is entirely protected against counterfeiting. (Dutta et al.2020)

7.1. Importance of Supply Chain Transparency

In this chapter, the intersection of cloud computing and blockchain technologies from a use case perspective is researched. The open-source technology stack is explored as an enabler for the next generation of decentralized applications. Examples of three such supply chain use cases are provided. The importance of true transparency in the supply chain is introduced.

In all businesses, various aspects like the cost of operations, energy, and other inputs required for operation have become increasingly important topics on the agenda. Ideas like corporate social

responsibility are gaining importance. The need to deliver on the promises made to customers means that today companies not only need to think in terms of near real-time changes to the supply chain but also need to be able to communicate these changes and adjustments with customers. They need to have the capability to share the truth of what is happening from the raw material processor to the retailer. Customers should not only be able to verify the authenticity of the various products irrespective of their location but also be able to access all the facets of the journey of that product. Providing proof of being counterfeit-free and conveying all the information, irrespective of the number of nodes or processors involved in the lifecycle of the product, is essential. Cloud and edge computing solutions become enablers in the journey of this information and transparency, along with the combination of blockchain technology.

7.2. Benefits of Blockchain in Supply Chain Tracking

The supply chain is an obvious and early target for blockchain. This is due to a reality that knits modern global commerce together, one that aspiring disruptors need to appreciate: few, if any, single companies still truly control an entire supply chain. Even large corporations enroll third parties, a mix of suppliers, manufacturers, transportation and logistics firms, and others to move component parts and products from origin to destination. Blockchain can enable a level of decentralized trust and coordination needed for optimizing the supply chain. By democratizing access to historical trade and logistics recordings across companies and regions, enterprises can integrate the logistics processes of all the parties involved on a single platform and still manage assured access control over who can interact and access data.

Additionally, with immutability in trade and logistics record keeping powered by decentralized trust, both the enterprises and the individual parties can secure greater transparency in real time, resulting in reduced fraudulent activities through sharing of consignment tracking and monitoring, and in improved customer satisfaction through trusted verified delivery estimations. Finally, with automatic verification at every stage enabled by integrating smart contracts into the supply chain process flow ledgers, enterprises can increase the speed of goods transit, reduce discrepancies, and enforce contractual compliance, resulting in reduced structural delays and human errors. (Udeh et al., 2024)

7.3. Real-world Examples

The concept of a decentralized application (dApp) is a primary blockchain use case. Instead of running on a single computer, a decentralized application operates on a P2P network of computers. These types of dApps enable developers to leverage features of blockchain ledgers to create consumer and/or

enterprise applications, where users maintain full control over their data and where they are incentivized and rewarded for their contribution to and use of the network.

Examples of dApps can be found in different fields, such as in the area of finance: a prediction market platform; a decentralized autonomous organization; and a money market protocol. In the area of social networks, there are various platforms. A secure multi-blockchain and multi-currency wallet for users is also available. The Ethereum network has the most dApps. There are a decentralized network of nodes that are able to communicate with one another; a blockchain game where players can buy, breed, and sell digital cats; and a cryptocurrency wallet. It is often used to store Ethereum and ERC20 tokens. There are many dApp marketplaces, from various app stores to others that list hundreds of applications grouped by the network they operate on.

8. Digital Identity Verification using Blockchain

The importance of digital identity verification has grown in recent years as online and e-commerce activities continue to proliferate. More and more services take advantage of digital apps to enable various activities and transactions, creating new digital personas users identify with. Secure digital identity verification becomes significant to control the digital crime issue. Blockchain technology becomes a major facilitator for protecting digital identities from being stolen.

The idea of implementing a blockchain-based identity verification system has been proposed. The approach of injecting digital identity data into blockchain nodes is needed. The strong linkage between digital tokens and human process entities has been established. Then digital identity verification can be performed by retrieving missing information. The idea of merging miners who are working both in blockchain mining and in the exchange of private information check processes has been proposed. This approach helps to verify the identity of the information receiver and prevents data leakage. In addition, attempts to index the user's private data still allow quick searches but do not access the data to be obtained, even removing the storage facility. Several companies have already used blockchain technology to protect digital identities. A software company has developed a blockchain-based digital identity verification application for establishing secure transactions. The company provides a blockchain-based platform for a variety of applications in multiple industries. The security and convenience of the program have been enhanced to protect the identity and data of both legal entities and individual customers. At the same time, a product is created that allows users to record their marriages, diplomas, and birth certificates using the blockchain platform. This effort is aimed at solving

the problem of legal identity in the world. Once the identity of a person is misused in these countries, it is impossible to register and verify through all centralized systems. (Zwitter et al., 2020)

8.1. Challenges in Traditional Identity Verification

Identity theft, as well as data breaches from service-providing organizations, has led to a loss of trust in online transactions. Malicious users with fake identities take advantage of the fact that governments have not been able to provide efficient online identity solutions for the majority of their citizens. In some countries, the only online government service that uses a person's real identity is electronically filed tax returns. Critics argue that later down the line, hidden costs can be reflected in a trade-off between identity confirmation assurance levels and the strength of the protections developed to secure identified transactions. Service providers following the principle of risk-based authentication assign different risks to different types of transactions. For organizations, these assignments are based on considering integrity and confidentiality measures along with continually ensuring the security of their applications.

The problem with risk-based user authentication is that the privacy of the user is not fully guaranteed. Privacy and identity management issues have been defined as the root of online insecurity. In Europe, the General Data Protection Regulation clarifies that any processing of personal data has to be justified under one of six lawful conditions, determined by the Data Protection Act. The ePrivacy regulation made it impossible to delegate authentication services to companies not registered inside European borders. However, two years later, the European Commission included an article on eID to be one of the priorities of the eEurope action plan. Since then, the European Commission has spent millions in order to support the mutual recognition by Member States of eIDs. Evidently, more efforts are necessary in order to establish an effective eID solution supporting an internal European digital single market, compatible with international regulations, and at the same time, acceptable to citizens. (Yeung & Bygrave, 2022)

8.2. Advantages of Blockchain-based Identity Verification

Decentralized identity verification systems offer substantial potential advantages over their centralized counterparts. It is worth noting that the capacity to assure that an individual is who he or she claims to be is foundational for many real-world interactions. Some examples include proof of age and proof of citizenship, as well as essential financial services such as obtaining a loan, receiving unemployment or other benefits to which an individual may be entitled, or even downloading free digital content. Despite

this need, on a global basis, there are believed to be over one billion people who lack the conventional forms of official ID which are often required for these sorts of transactions. (Glöckler et al.2024)

8.3. Case Studies

Embarking on the journey of bridging two major distinct technologies: cloud computing and blockchain, we develop several case studies or use cases for exploring our propositions. Some of the use cases are realistic where blockchain can be replaced by cloud and still work, but blockchain offers different advantages than cloud does at a significantly low cost. Some use case scenarios are hypothetical, but may serve to enable new applications or technologies when cloud meets blockchain. In our use cases, we take examples from decentralized applications, supply chain tracking, counterfeit warranties app, digital identity and access management, decentralized VPN, and decentralized AI marketplaces.

For the use case of decentralized applications, we consider the realistic scenario of cloud and blockchain acting as generalized backend services. The use case assumes that the public blockchain platforms over the cloud host smart contract software comprising data and logic for an application service. In this model, the overall client payment and computation retrieval architecture are such that when trusted code is uploaded, it can be programmed to enable the retrieval requests. These requests can return the response required by a phase two operation on the relied cloud resource. For decentralized cloud services, we consider the hypothetical scenario where a cost-effective architecture is built using both cloud and blockchain that can support storage and computation, enabling the deployment of cloud services to end users. The decentralized cloud services employ an off-chain cloud database managed by cloud service providers operating in a decentralized setup on a large scale. These service providers may offer cost-effective, untrusted resources.

10. Conclusion

Blockchain and cloud computing are technologies which by their very nature are designed to enable collaboration and cryptocurrency paradigm. However, as we have explored the various use cases, it has become clear how cloud computing can address some of the inherent challenges with blockchain implementation, making it more viable for a range of different applications. We have seen that blockchain is a fit for a range of applications given its key attributes of being trust, transparency, and auditability with its goal to enable decentralization. However, it is also clear at the same time that cloud computing is also a good for a wide range of applications, because of the need for centralized management and control for enabling business logic that blockchain doesn't necessarily provide. It is

important to stress that cloud computing partnering with Blockchain technology can enhance the key attributes of blockchain, and bring decentralization to a wide range of different applications. Blockchain teamed up with cloud computing is a secure, flexible, and cost-efficient solution. Given the rapid grow of IoT devices and platforms, and the growth of decentralized and centralized business applications, the work put forth in this chapter will be useful for developers, researchers, and decision makers on the relationship and trade-offs between cloud and blockchain technologies, and in assisting in trying to decide which technologies may suit an application best. In particular, the analysis framework, and the case studies, user can see that many centralized enterprise applications can be replaced by decentralized ones.

References

- GEORGE, D. A. S., & GEORGE, A. H. (2022). Potential risk: Hosting cloud services outside the country. *International Journal of Advanced Research in Computer and Communication Engineering*, 11(4), 5-11. [researchgate.net](https://www.researchgate.net)
- Gietzmann, M. & Grossetti, F. (2021). Blockchain and other distributed ledger technologies: where is the accounting?. *Journal of Accounting and Public Policy*. ssrn.com
- Miyachi, K. & Mackey, T. K. (2021). hOCBS: A privacy-preserving blockchain framework for healthcare data leveraging an on-chain and off-chain system design. *Information processing & management*. cafetarjome.com
- Dong, S., Abbas, K., Li, M., & Kamruzzaman, J. (2023). Blockchain technology and application: an overview. *PeerJ Computer Science*. peerj.com
- Sovacool, B. K., Upham, P., & Monyei, C. G. (2022). The “whole systems” energy sustainability of digitalization: Humanizing the community risks and benefits of Nordic datacenter development. *Energy Research & Social Science*. sciencedirect.com
- Qian, X., & Papadonikolaki, E. (2021). Shifting trust in construction supply chains through blockchain technology. *Engineering, Construction and Architectural Management*, 28(2), 584-602. ucl.ac.uk
- Murthy, C. V. N. U. B., Shri, M. L., Kadry, S., & Lim, S. (2020). Blockchain based cloud computing: Architecture and research challenges. *IEEE access*. iee.org
- Gonczol, P., Katsikouli, P., Herskind, L., & Dragoni, N. (2020). Blockchain implementations and use cases for supply chains-a survey. *Ieee Access*. iee.org
- Habib, G., Sharma, S., Ibrahim, S., Ahmad, I., Qureshi, S., & Ishfaq, M. (2022). Blockchain technology: benefits, challenges, applications, and integration of blockchain technology with cloud computing. *Future Internet*, 14(11), 341. mdpi.com

- Teymourifar, A., Rodrigues, A. M., & Ferreira, J. S. (2021). A comparison between simultaneous and hierarchical approaches to solve a multi-objective location-routing problem. *Graphs and Combinatorial Optimization: from Theory to Applications: CTW2020 Proceedings*, 251-263. [\[HTML\]](#)
- Mio, C., Panfilo, S., & Blundo, B. (2020). Sustainable development goals and the strategic role of business: A systematic literature review. *Business strategy and the environment*, 29(8), 3220-3245. [unive.it](#)
- Popchev, I. & Radeva, I. (2024). Decentralized Application (dApp) Development and Implementation. *Cybernetics and Information Technologies*. [sciendo.com](#)
- Santana, C. & Albareda, L. (2022). Blockchain and the emergence of Decentralized Autonomous Organizations (DAOs): An integrative model and research agenda. *Technological Forecasting and Social Change*. [sciencedirect.com](#)
- Zhaobin, C., Jingrong, H., Chenyang, F., & Xinyi, J. (2024). Research on smart contract and front-end technology integration in Dapp development. *Academic Journal of Computing & Information Science*, 7(3), 55-62. [francis-press.com](#)
- Jamil, F., Iqbal, N., Ahmad, S., & Kim, D. (2021). Peer-to-peer energy trading mechanism based on blockchain and machine learning for sustainable electrical power supply in smart grid. *Ieee Access*. [ieee.org](#)
- Dutta, P., Choi, T. M., Somani, S., & Butala, R. (2020). Blockchain technology in supply chain operations: Applications, challenges and research opportunities. *Transportation research part e: Logistics and transportation review*, 142, 102067. [nih.gov](#)
- Udeh, E. O., Amajuoyi, P., Adeusi, K. B., & Scott, A. O. (2024). The role of IoT in boosting supply chain transparency and efficiency. [tgdaddy.com](#)
- Zwitter, A. J., Gstrein, O. J., & Yap, E. (2020). Digital identity and the blockchain: universal identity management and the concept of the “Self-Sovereign” individual. *Frontiers in Blockchain*. [frontiersin.org](#)
- Yeung, K. & Bygrave, L. A. (2022). Demystifying the modernized European data protection regime: Cross-disciplinary insights from legal and regulatory governance scholarship. *Regulation & Governance*. [wiley.com](#)
- Glöckler, J., Sedlmeir, J., Frank, M., & Fridgen, G. (2024). A systematic review of identity and access management requirements in enterprises and potential contributions of self-sovereign identity. *Business & Information Systems Engineering*, 66(4), 421-440. [springer.com](#)

A new Survey on Metaheuristic Algorithms for Solving the Optimization Problems

Assistant Prof. Samera Khaleel Ibrahim

Assistant Prof. Dr. Zainib Hatif Abbas

Department of Statistic
College of Administration and Economics
University of Baghdad, Iraq

Institute of Genetic Engineering and
Biotechnology for Post Graduate Studies

Samera.khaleel@coadec.uobaghdad.edu.iq
Zainab.hatif@ige.uobaghdad.edu.iq

A new Survey on Metaheuristic Algorithms for Solving the Optimization Problems

Assistant Prof. Samera Khaleel Ibrahim

Department of Statistic
College of Administration and Economics

University of Baghdad, Iraq

Zainab.hatif@ige.uobaghdad.edu.iqSamera.khaleel@coadec.uobaghdad.edu.iq

Assistant Prof. Dr. Zainib Hatif Abbas

Institute of Genetic Engineering and
Biotechnology for Post Graduate Studies

Abstract

Metaheuristic algorithms are nature-inspired, have been a high level of search techniques, designed for global search to efficiently large solution spaces to find optimal or near-optimal solutions unlike traditional algorithms. These algorithms strike a balance between a suitable solution and its search time. The appropriate solution is selected from a group of solutions during several cycles by applying certain rules or criteria. Exploration and Exploitation are the two common features used to solve the problem in any optimization method. Exploration is the stage of searching for solutions within the search space by expanding this field to unexplored areas, and exploitation focuses on search areas that may contain solutions close to optimal solutions. Metaheuristic algorithms have been recently spread to solve many complex problems in different fields due to their flexibility and simplicity. In this paper, the metaheuristic algorithms that have been designed during the last ten years are presented and classified according to their characteristics and fields of use.

Keyword: Metaheuristic Algorithms, Optimization Problems

1. Introduction

Metaheuristic algorithms are optimization techniques that are designed to find an adequate solution for a broad range of optimization problems. These algorithms stand out from other optimization techniques in several ways. Firstly, they are derivative-free, meaning that they do not require any sort of calculation of derivatives in the search space, as opposed to gradient-based search techniques. This makes metaheuristic algorithms much simpler, more flexible, and more capable of avoiding local optima, making them highly effective for handling challenging optimization tasks. The stochastic nature is another characteristic of metaheuristic algorithms, which implies they begin the optimization process by generating random results. This makes it more likely that the algorithms will be able to avoid premature convergence and quickly and effectively examine the search space. Metaheuristics balance between exploration and profit to accomplish this. During the discovery step, the algorithms

thoroughly examine the search space's interesting regions, and then, in the analysis phase, they carry out local searches in these regions to perceive the most gilt-edge resolution. The specific and primary key advantages of metaheuristic algorithms are their versatility and flexibility. They can be modified easily to fit the specific requirements of a particular problem, making them an ideal solution for a broad range of optimization problems across various fields of engineering and science. metaheuristics have been successfully applied in electrical engineering optimization, in industrial scheduling and transportation, in civil engineering for bridge and building design, networking, in data mining for classification, prediction, clustering, and system modeling. Metaheuristics are a powerful and widely used framework for solving optimization problems. They provide a set of guidelines and strategies that can be used to develop efficient heuristic optimization algorithms. They are employed in both mathematical optimization and computer science. Metaheuristics enable the efficient exploration of a large search space by testing a subdivision containing elucidations that could ordinarily be excessively sizeable to be wholly recapitulated or explored. In this paper, the metaheuristic algorithms that were designed in the period between (2014-2024) are presented, defined and classified.

2. Meta-heuristic Algorithms

In this section, presented some of metaheuristic algorithms are designed in the last ten years.

2.1 Grey Wolf Optimizer (GWO).

GWO inspired by grey wolves, mimics the leadership hierarchy and hunting mechanism of grey wolves in nature. Four types of grey wolves such as alpha, beta, delta, and omega are employed for simulating the leadership hierarchy. In addition, the three main steps of hunting, searching for prey, encircling prey, and attacking prey, are implemented. Grey wolves are considered as apex predators, meaning that they are at the top of the food chain. Grey wolves mostly prefer to live in a pack. The group size is 5–12 on average. Of particular interest is that they have a very strict social dominant hierarchy. The leaders are a male and a female, called alphas. The alpha is mostly responsible for making decisions about hunting, sleeping place, time to wake, and so on. The alpha's decisions are dictated to the pack. However, some kind of democratic behavior has also been observed, in which an alpha follows the other wolves in the pack. In gatherings, the entire pack acknowledges the alpha by holding their tails down. The alpha wolf is also called the dominant wolf since his/her orders should be followed by the pack. The alpha wolves are only allowed to mate in the pack. Interestingly, the alpha is not necessarily the strongest member of the pack but the best in terms of managing the pack. This shows that the organization and discipline of a pack is much more important than its strength. The second level in the hierarchy of grey wolves is beta. The betas are subordinate wolves that help the alpha in decision-making or other pack activities. The beta wolf can be either male or female, and he or

she is probably the best candidate to be the alpha. In case one of the alpha wolves passes away or becomes very old. The beta wolf should respect the alpha, but commands the other lower-level wolves as well. It plays the role of an advisor to the alpha and discipliner for the pack. The beta reinforces the alpha's commands throughout the pack and gives feedback to the alpha. The lowest ranking grey wolf is omega. The omega plays the role of scapegoat. Omega wolves always have to submit to all the other dominant wolves. They are the last wolves that are allowed to eat. It may seem the omega is not an important individual in the pack, but it has been observed that the whole pack face internal fighting and problems in case of losing the omega. This is due to the venting of violence and frustration of all wolves by the omega. This assists satisfying the entire pack and maintaining the dominance structure. In some cases, the omega is also the babysitters in the pack. If a wolf is not an alpha, beta, or omega, he/she is called subordinate (or delta in some references). Delta wolves have to submit to alphas and betas, but they dominate the omega. Scouts, sentinels, elders, hunters, and caretakers belong to this category. Scouts are responsible for watching the boundaries of the territory and warning the pack in case of any danger. Sentinels protect and guarantee the safety of the pack. Elders are the experienced wolves who used to be alpha or beta. Hunters help the alphas and betas when hunting prey and providing food for the pack. Finally, the caretakers are responsible for caring for the weak, ill, and wounded wolves in the pack. In addition to the social hierarchy of wolves, group hunting is another interesting social behavior of grey wolves. The main phases of grey wolf hunting are as follows:

- Tracking, chasing, and approaching the prey.
- Pursuing, encircling, and harassing the prey until it stops moving.
- Attack towards the prey.

In this hunting technique and the social hierarchy of grey wolves are mathematically modeled in order to design GWO and perform optimization. In this algorithm, the mathematical model consists of (social hierarchy) tracking, encircling, and attacking prey (Mirjalili et al., 2014).

2.2 Whale Optimization Algorithm (WOA)

Whale Optimization Algorithm (WOA) is a nature-inspired meta-heuristic optimization algorithm, mimics the social behavior of humpback whales. The algorithm is inspired by the bubble-net hunting strategy, designed by Mirjalili & Lewis in 2015. Whales are fancy creatures. They are considered as the biggest mammals in the world. An adult whale can grow up to 30 m long and 180 t weight. There are 7 different main species of this giant mammal such as killer, Minke, Sei, humpback, right, finback, and blue. Whales are mostly considered as predators. They never sleep because they have to breathe from the surface of oceans. In fact, half of the brain only sleeps. The interesting thing about the whales is that they are considered as highly intelligent animals with emotion. Whales have common cells in certain

areas of their brains similar to those of human called spindle cells. These cells are responsible for judgment, emotions, and social behaviors in humans. In other words, the spindle cells make us distinct from other creatures. Whales have twice number of these cells than an adult human which is the main cause of their smartness. It has been proven that whale can think, learn, judge, communicate, and become even emotional as a human does, but obviously with a much lower level of smartness. It has been observed that whales (mostly killer whales) are able to develop their own dialect as well. Another interesting point is the social behavior of whales. They live alone or in groups. However, they are mostly observed in groups. Some of their species (killer whales for instance) can live in a family over their entire life period. One of the biggest baleen whales is humpback whales (*Megaptera novaeangliae*). An adult humpback whale is almost as size of a school bus. Their favorite prey are krill and small fish herds. The most interesting thing about the humpback whales is their special hunting method. This foraging behavior is called bubble-net feeding method. Humpback whales prefer to hunt school of krill or small fishes close to the surface. this behavior was only investigated based on the observation from surface. However, investigated this behavior utilizing tag sensors. They captured 300 tag-derived bubble-net feeding events of 9 individual humpback whales. They found two maneuvers associated with bubble and named them ‘upward-spirals’ and ‘double-loops’. In the former maneuver, humpback whales dive around 12 m down and then start to create bubble in a spiral shape around the prey and swim up toward the surface. The later maneuver includes three different stages: coral loop, lobtail, and capture loop. It is worth mentioning here that bubble-net feeding is a unique behavior that can only be observed in humpback whales. In this work the spiral bubble-net feeding maneuver is mathematically modeled in order to perform optimization.

The mathematical model of the WOA consists of three parts; encircling prey, Humpback whales can recognize the location of prey and encircle them. Since the position of the optimal design in the search space is not known a priori, the WOA algorithm assumes that the current best candidate solution is the target prey or is close to the optimum. After the best search agent is defined, the other search agents will hence try to update their positions towards the best search agent. the humpback whales also attack the prey with the bubble-net strategy (exploitation phase) In order to mathematically model the bubble-net behavior of humpback whales, two approaches are designed are (Shrinking encircling mechanism, and Spiral updating position). Search for prey (exploration phase), humpback whales search randomly according to the position of each other. to force search agent to move far away from a reference whale. In contrast to the exploitation phase, we update the position of a search agent in the exploration phase according to a randomly chosen search agent instead of the best search agent found so far. This

mechanism and emphasize exploration and allow the WOA algorithm to perform a global search (Mirjalili & Lewis, 2016).

2.3 Moth-Flame Optimization (MFO)

MFO algorithm nature-inspired optimization paradigm designed by Mirjalili in 2015. The main inspiration of this optimizer is the navigation method of moths in nature called transverse orientation. Moths fly in night by maintaining a fixed angle with respect to the moon, a very effective mechanism for travelling in a straight line for long distances. However, these fancy insects are trapped in a deadly spiral path around artificial lights. Since the moon is far away from the moth, this mechanism guarantees flying in straight line. The same navigation method can be done by humans. Suppose that the moon is in the south side of the sky and a human wants to go the east. If he keeps moon of his left side when walking, he would be able to move towards the east on a straight line. Despite the effectiveness of transverse orientation, we usually observe that moths fly spirally around the lights. In fact, moths are tricked by artificial lights and show such behaviors. This is due to the inefficiency of the transverse orientation, in which it is only helpful for moving in straight line when the light source is very far. When moths see a human-made artificial light, they try to maintain a similar angle with the light to fly in straight line. It that the moth eventually converges towards the light. This behavior is modeled mathematically to Moth-Flame Optimization (MFO) algorithm.

In the MFO algorithm, it is assumed that the candidate solutions are moths and the problem's variables are the position of moths in the space. Therefore, the moths can fly in 1-D, 2-D, 3-D, or hyper dimensional space with changing their position vectors. Since the MFO algorithm is a population-based algorithm, the set of moths is represented in a matrix. For all the moths, that there is an array for storing the corresponding fitness values. Key components in the algorithm are flames and moths. It should be noted here that moths and flames are both solutions. The difference between them is the way we treat and update them in each iteration. The moths are actual search agents that move around the search space, whereas flames are the best position of moths that obtains so far. In other words, flames can be considered as flags or pins that are dropped by moths when searching the search space. Therefore, each moth searches around a flag (flame) and updates it in case of finding a better solution. With this mechanism, a moth never loses its best solution. The MFO algorithm is a three-tuple that approximates the global optimal of the optimization problems, so it potentially able to improve the initial random solutions and convergence to a better point in the search space (Mirjalili, 2015).

2.4 Crow Search Algorithm (CSA)

CSA based on the intelligent behavior of crows. It is a population-based technique which works based on this idea that crows store their excess food in hiding places and retrieve it when the food is

needed. Crows (crow family or corvids) are considered the most intelligent birds. They contain the largest brain relative to their body size. Based on a brain-to-body ratio, their brain is slightly lower than a human brain. Evidences of the cleverness of crows are plentiful. They have demonstrated self-awareness in mirror tests and have tool-making ability. Crows can remember faces and warn each other when an unfriendly one approaches. Moreover, they can use tools, communicate in sophisticated ways and recall their food's hiding place up to several months later. Crows have been known to watch other birds, observe where the other birds hide their food, and steal it once the owner leaves. If a crow has committed thievery, it will take extra precautions such as moving hiding places to avoid being a future victim. In fact, they use their own experience of having been a thief to predict the behavior of a pilferer, and can determine the safest course to protect their caches from being pilfered. based on intelligent behaviors, a population-based metaheuristic algorithm, CSA, is developed. It is assumed that there is a d -dimensional environment including a number of crows. The number of crows (flock size) is N and the position of crow i at time (iteration) in the search space is specified by a vector $iter_{max}$ is the maximum number of iterations. Each crow has a memory in which the position of its hiding place is memorized. At iteration, the position of hiding place of crow i . This is the best position that crow i has obtained so far. In memory of each crow the position of its best experience has been memorized. Crows move in the environment and search for better food sources (hiding places). Assume that at iteration, crow j wants to visit its hiding place, m_j ; iter. At this iteration, crow i decides to follow crow j to approach to the hiding place of crow j . In this case, two states may happen:

State 1: Crow j does not know that crow i is following it. As a result, crow i will approach to the hiding place of crow j . In this case, the new position of crow i .

State 2: Crow j knows that crow i is following it. As a result, in order to protect its cache from being pilfered, crow j will fool crow i by going to another position of the search space. Metaheuristic algorithms should provide a good balance between diversification and intensification. In CSA, intensification and diversification are mainly controlled by the parameter of awareness probability (AP). By decrease of the awareness probability value, CSA tends to conduct the search on a local region where a current good solution is found in this region. As a result, using small values of AP, increases intensification. On the other hand, by increase of the awareness probability value, the probability of searching the vicinity of current good solutions decreases and CSA tends to explore the search space on a global scale (randomization). As a result, use of large values of AP increases diversification (Askarzadeh, 2016).

2.5 Dragonfly Algorithm (DA).

DA is swarm intelligence optimization technique. The main inspiration of the DA algorithm originates from the static and dynamic swarming behaviors of dragonflies in nature. Two essential phases of optimization, exploration and exploitation, are designed by modelling the social interaction of dragonflies in navigating, searching for foods, and avoiding enemies when swarming dynamically or statistically. this algorithm is able to improve the initial random population for a given problem, converge towards the global optimum. Dragonflies (Odonata) are fancy insects. There are nearly 3000 different species of this insect around the world. main milestones: nymph and adult. They spend the major portion of their lifespan in nymph, and they undergo metamorphism to become adult. Dragonflies are considered as small predators that hunt almost all other small insects in nature. Nymph dragonflies also predate on other marine insects and even small fishes. The interesting fact about dragonflies is their unique and rare swarming behavior. Dragonflies swarm for only two purposes: hunting and migration. The former is called static (feeding) swarm, and the latter is called dynamic (migratory) swarm. In static swarm, dragonflies make small groups and fly back and forth over a small area to hunt other flying preys such as butterflies and mosquitoes. Local movements and abrupt changes in the flying path are the main characteristics of a static swarm. In dynamic swarms, however, a massive number of dragonflies make the swarm for migrating in one direction over long distances. The main inspiration of the DA algorithm originates from static and dynamic swarming behaviors. These two swarming behaviors are very similar to the two main phases of optimization using meta-heuristics: exploration and exploitation. Dragonflies create sub-swarms and fly over different areas in a static swarm, which is the main objective of the exploration phase. In the static swarm, however, dragonflies fly in bigger swarms and along one direction, which is favorable in the exploitation phase. These two phases are mathematically implemented. Operators for exploration and exploitation, the behavior of swarms follows three primitive principles.

- Separation, which refers to the static collision avoidance of the individuals from other individuals in the neighborhood.
- Alignment, which indicates velocity matching of individuals to that of other individuals in neighborhood.
- Cohesion, which refers to the tendency of individuals towards the center of the mass of the neighborhood.

The main objective of any swarm is survival, so all of the individuals should be attracted towards food sources and distracted outward enemies (Mirjalili, 2016).

2.6 Emperor Penguin Optimizer (EPO)

EPO, which mimics the huddling behavior of emperor penguins. The emperor penguin, scientifically named as *Aptenodytes forsteri*, is the tallest and heaviest in all of the penguin species. The male and female emperor penguins are very much similar in terms of plumage and size. The dorsal side and head are black with white belly, pale yellow breast, and bright yellow ear patches. Emperor penguins spend their lives in open ice and breeds during winter season. During the breeding season, emperor penguins come ashore in huge colonies which includes hundreds of thousands of emperor penguins. Females lay a single egg and can travel 50 miles to reach the ocean for hunting. The emperor penguin is a social like animal by its foraging behavior and hunting together in a group. In sea, emperor penguins can dive up to 1900 feet deeper and stay under sea for more than 25 min. Emperor penguins are flightless like other penguins' species with stiffened and flattened wings. Emperor penguins are the only species that huddles to survive during the Antarctic winter. The huddling behavior of emperor penguins is decomposed into four phases:

- Generate and determine the huddle boundary of emperor penguins.
- Calculate the temperature profile around the huddle.
- Determine the distance between emperor penguins.
- Relocate the effective mover.

the huddling behavior of emperor penguins is mathematically modeled. The main aim of this model is to find an effective mover. The huddle is assumed to be situated on two-dimensional L-shape polygon plane. Firstly, emperor penguins generate the huddle boundary randomly. Thereafter, the temperature profile around the huddle is computed. The distance between emperor penguins is also calculated which will be helpful for more exploration and exploitation. Finally, the effective mover i.e., the best optimal solution is obtained and recompute the boundary of huddle with updated positions of emperor penguins (or search agents) (Dhiman & Kumar, 2018).

2.7 Political Optimizer (PO)

PO inspired by the multi-phased process of politics. It is the mathematical mapping of all the major phases of politics such as constituency allocation, party switching, election campaign, inter-party election, and parliamentary affairs. PO algorithm assigns each solution a dual role by logically dividing the population into political parties and constituencies, which facilitates each candidate to update its position with respect to the party leader and the constituency winner. Moreover, a novel position updating strategy called recent past-based position updating strategy (RPPUS) is introduced, which is the mathematical modeling of the learning behaviors of the politicians from the previous election. Politics itself is a process of optimization from two perspectives: each individual optimizes its goodwill

to win the election and each party tries to maximize its number of seats in parliament to form a government. These aspects make politics an ideal inspiration for an optimization algorithm because an individual (a party member) may be considered a candidate solution, individual's goodwill is considered the position of the candidate solution in the search space, and goodwill of a political member can be defined by many performances related parameters which can be mimicked by design variables or components of the position vector of a candidate solution. Election may be considered the evaluation (objective) function and the number of votes obtained by an individual in election is mapped by the fitness of the candidate solution (Gutekunst & Williamson, 2020).

2.8 Poor and Rich Optimization (PRO).

This algorithm is inspired by the efforts of the two groups of the poor and the rich to achieve wealth and improve their economic situation. The rich are always trying to increase their class gap with the poor by acquiring wealth from different ways. On the other hand, the poor try to gain wealth and reduce their class gap by modeling the rich. This struggle is always going on and should be mention that the poor may get rich and vice versa. PRO is designed based on a real social phenomenon which can be considered as a solution for complicated optimization problems. In fact, the main idea of the proposed algorithm is summarized in two cases:

1. Each member of the poor group tries to improve their status and reduce their class gap by learning from the rich.
2. Each member of the rich group tries to increase the class gap.

The first population including the rich and poor was randomly designed as a uniform vector and this population will be arranged based on the objective function. The first part of this population including the best values will be considered for the rich group and the second part will be defined for the poor group (Samareh Moosavi & Bardsiri, 2019).

2.9 Hunger Games Search (HGS)

HGS is designed according to the hunger-driven activities and behavioral choice of animals. This dynamic, fitness-wise search method follows a simple concept of "Hunger" as the most crucial homeostatic motivation and reason for behaviors, decisions, and actions in the life of all animals to make the process of optimization more understandable and consistent for new users and decision-makers. The Hunger Games Search incorporates the concept of hunger into the feature process; in other words, an adaptive weight based on the concept of hunger is designed and employed to simulate the effect of hunger on each search step. It follows the computationally logical rules (games) utilized by almost all animals and these rival activities and games are often adaptive evolutionary by securing higher chances of survival and food acquisition. This method's main feature is its dynamic nature,

simple structure, and high performance in terms of convergence and acceptable quality of solutions (Nd, 2021).

2.10 Henry gas solubility optimization (HGSO)

HGSO mimics the behavior governed by Henry's law to solve challenging optimization problems. Henry's law is an essential gas law relating the amount of a given gas that is dissolved to a given type and volume of liquid at a fixed temperature. The HGSO algorithm imitates the huddling behavior of gas to balance exploitation and exploration in the search space and avoid local optima. HGSO is a designed population-based metaheuristic optimization algorithm based on principles of physics. It is that the ability to balance exploration and exploitation plays a significant role in this algorithm. This characteristic makes HGSO suitable for solving complex optimization problems with many local optimal solutions because it keeps a population of solutions and investigates a large area to find the best global solution (Hashim et al., 2019).

2.11 Archimedes optimization algorithm (AOA).

AOA is a meta-heuristic algorithm designed to solve the optimization problems. It is devised with inspirations from an interesting law of physics Archimedes' Principle. It imitates the principle of buoyant force exerted upward on an object, partially or fully immersed in fluid, is proportional to weight of the displaced fluid. It is a high-performance optimization tool with respect to convergence speed and exploration-exploitation balance, as it is effectively applicable for solving complex problems (Hashim et al., 2021).

2.12 Ebola optimization search algorithm (EOSA).

EOSA is a nature-inspired and biology-based algorithm that was inspired by the natural and biological process of Ebola. It has been formulated the propagation of the disease using a mathematical and a susceptible infected and recovered (SIR) model. Thereafter, an algorithmic design of the procedure for the optimization process was done (Ovelade & Ezugwu, 2021).

2.13 War Strategy Optimization (WSO).

WSO is a metaheuristic optimization algorithm based on ancient war strategy, so on the strategic movement of army troops during the war. War Strategy is modeled as an optimization process wherein each soldier dynamically moves towards the optimum value. The algorithm models two popular war strategies, attack and defense strategies. The positions of soldiers on the battlefield are updated in accordance with the strategy implemented. To improve the algorithm's convergence and robustness, a novel weight updating mechanism and a weak soldier's relocation strategy are introduced. The war strategy algorithm achieves good balance of the exploration and exploitation stages (Ayyarao et al., 2022).

2.14 The Artificial Electric Field Algorithm (AEFA)

AEFA a physics-inspired metaheuristic, drawing inspiration from Coulomb's law and electrostatic force; however, while AEFA has demonstrated efficacy, it can face challenges such as convergence issues and suboptimal solutions, especially in high-dimensional problems. To overcome these challenges, employed to add more diversity. The integration of these mechanisms in AEFA aims to expand its search space, enhance exploration potential, avoid local optima, and achieve improved performance, robustness, and a more equitable equilibrium between local intensification and global diversification(Hussien et al., 2024).

3. Meta-heuristic Algorithms Classification

The use of advanced technology in various fields of science is increasing the complexity of the problems to be solved. The shortcomings of traditional optimization techniques resulted in the emergence of the metaheuristic optimization algorithm for solving complex engineering problems. As a result, new optimization algorithms become a ray of hope. Meta-heuristic methods: Meta-heuristics methods are considered to be global best optimization algorithms and possess several advantages, such as robustness, performance reliability, simplicity, ease of implementation.

Meta-heuristic algorithms have been classified, such as:

- a. Evolutionary-based algorithms: These algorithms are originated from the theory of evolution, such as: genetic algorithm (GA) (Whitley, 1994), Differential Evolution (DE), (Storn & Price, 1997), Evolutionary Strategy (ES) (Simon, 2013).
- b. Swarm-based algorithms: These algorithms emulate the social behavior and the collective decision-making of various social groups. In these algorithms, the explanation for reaching a given objective is usually based on bio-community intelligence and collective action such as: GWO, WOA, MFO, CSA, DA, EOSA, EPO.
- c. Physics-based algorithms: The physics-based algorithms have been influenced by the laws of natural physics such as: AEFA, AOA, HGSO.
- d. Human behavior-based algorithms: optimization algorithms inspired by human beings' social behavior such as: WSO, HGS, PRO, PO.

4.Concolusion

This work surveyed several important metaheuristic methods as they are described for last ten years. Some of them are inspired by the social behavior of insects and animals, some are inspired by human behavior, and some are inspired by physical and biological laws. Since the 1950s, with the development

of computer science, metaheuristic algorithms inspired by nature have been designed to solve and improve complex problems that cannot be solved by traditional methods.

References

- Askarzadeh, A. (2016). A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm. *Computers and Structures*, 169, 1–12. <https://doi.org/10.1016/j.compstruc.2016.03.001>
- Ayyarao, T. S. L. V, Ramakrishna, N. S. S., Elavarasan, R. M., Polumahanthi, N., Rambabu, M., Saini, G., Khan, B., & Alatas, B. (2022). War strategy optimization algorithm: a new effective metaheuristic algorithm for global optimization. *IEEE Access*, 10, 25073–25105.
- Dhiman, G., & Kumar, V. (2018). Emperor penguin optimizer: A bio-inspired algorithm for engineering problems. *Knowledge-Based Systems*, 159(June), 20–50. <https://doi.org/10.1016/j.knsys.2018.06.001>
- Gutekunst, S. C., & Williamson, D. P. (2020). rna Jou pro of. *Operations Research Letters*, 110645. <https://doi.org/10.1016/j.orl.2020.02.011>
- Hashim, F. A., Houssein, E. H., Mabrouk, M. S., Al-Atabany, W., & Mirjalili, S. (2019). Henry gas solubility optimization: A novel physics-based algorithm. *Future Generation Computer Systems*, 101, 646–667. <https://doi.org/10.1016/j.future.2019.07.015>
- Hashim, F. A., Hussain, K., Houssein, E. H., Mabrouk, M. S., & Al-Atabany, W. (2021). Archimedes optimization algorithm: a new metaheuristic algorithm for solving optimization problems. *Applied Intelligence*, 51(3), 1531–1551. <https://doi.org/10.1007/s10489-020-01893-z>
- Hussien, A. G., Pop, A., Kumar, S., Hashim, F. A., & Hu, G. (2024). A Novel Artificial Electric Field Algorithm for Solving Global Optimization and Real-World Engineering Problems. *Biomimetics*, 9(3), 1–37. <https://doi.org/10.3390/biomimetics9030186>
- Mirjalili, S. (2015). Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowledge-Based Systems*, 89, 228–249. <https://doi.org/10.1016/j.knsys.2015.07.006>
- Mirjalili, S. (2016). Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. *Neural Computing and Applications*, 27(4), 1053–1073. <https://doi.org/10.1007/s00521-015-1920-1>
- Mirjalili, S., & Lewis, A. (2016). The Whale Optimization Algorithm. *Advances in Engineering Software*, 95, 51–67. <https://doi.org/10.1016/j.advengsoft.2016.01.008>
- Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey Wolf Optimizer. *Advances in Engineering Software*, 69, 46–61. <https://doi.org/10.1016/j.advengsoft.2013.12.007>
- Nd, C.-. (2021). <http://creativecommons.org/licenses/by-nc-nd/4.0/>.
- Ovelade, O. N., & Ezugwu, A. E. (2021). Ebola Optimization Search Algorithm: A new nature-inspired metaheuristic algorithm for global optimization problems. *International Conference on Electrical, Computer, and Energy Technologies, ICECET 2021, March*. <https://doi.org/10.1109/ICECET52533.2021.9698813>
- Samareh Moosavi, S. H., & Bardsiri, V. K. (2019). Poor and rich optimization algorithm: A new human-based and multi populations algorithm. *Engineering Applications of Artificial Intelligence*, 86(November 2018), 165–181. <https://doi.org/10.1016/j.engappai.2019.08.025>
- Simon D., (2013),” *Evolutionary Optimization Algorithms*”. Hoboken, NJ, USA: Wiley.
- Storn R. & Price K., (1997), ``Differential evolution_A simple and efficient heuristic for

global optimization over continuous spaces," *J. Global Optim.*, vol. 11, no. 4, pp. 341_359.

Whitley D., (1994), ``A genetic algorithm tutorial," *Statist. Comput.*, vol. 4, no. 2, pp. 65_85.

Data Security Algorithm Using Magic Square With Order Five Based on GF(2⁸) Numbers

Ibrahim Malik ALattar¹ *, Abdul Monem S. Rahma²

¹, Department of Computer Science, University of Technology, Baghdad, Iraq.

**², Computer Science Department, college of Science ,Al-Maarif University,
Anbar, Iraq.**

ibrahiminter@yahoo.com¹

Monem.rahma@uoa.edu.iq²

Data Security Algorithm Using Magic Square With Order Five Based on $GF(2^8)$ Numbers

Ibrahim Malik ALattar¹ *, Abdul Monem S. Rahma²

¹, Department of Computer Science, University of Technology, Baghdad, Iraq.

², Computer Science Department, college of Science ,Al-Maarif University, Anbar, Iraq.

ibrahiminter@yahoo.com¹

Monem.rahma@uoa.edu.iq²

Abstract

Development this paper aims to develop a Cryptography algorithms based upon the advantages of magic squares to obtain high speed and high complexity. In the paper work on Magic Square of Order 5 was relied upon to generate an algorithm of encryption for the data specified with $GF(P)$ or $GF(2^8)$. The process done by put the key in the positions that agreed and put the message in the remaining positions at the Magic Square then the sum is found for each row, Column and diameter. The block encryption algorithm uses a block of size 10 and 14 element. The complexity obtained by the algorithm was $(256)^{10} \times (P)^{15}$ and $(256)^{14} \times (P)^{11}$ respectively, and the speed of each proposed algorithm was calculated and the results indicated that the first proposed method was better than the second method in terms of complexity, especially when using a high number prime in $GF(P)$ compared.

Keywords: Magic Square, Cryptography, linear equation system, finite field, $GF(2^8)$.

1. Introduction

The search for new cryptography techniques has led cryptologists and mathematicians to confiscate factor leading to success of the well-known essential attributes of the magic square. Magic Squares of order N (MS N) based on a number group that uses the numbers 1 to n^2 [1].

The story of magic square goes back to several centuries ago, which was discovered some of it emergence goes back to before birth in a lot of countries of world such as china , India and Middle east then moved to Europe through translation of Arabic resource then moved to the rest of the world[2].

Magic squares prosper and forward quickly with the emerging mathematics-based games like puzzles and Sudoku games, which have been especially appealing to the mathematicians [3], as we will see in this paper.

2. Related Work

This part provides an overview of the relevant literature on the various modifications to the uses of the magic square.

in the year 2009 [4], Ganapathy and Mani presented cryptography method based on the magic square and the RSA algorithm, encryption / decryption is based on the numbers that are generated by the magic square instead of ASCII values.

in 2014 [5], Researchers presented a technology in protecting data in cloud computing, using the RSA algorithm and magic square by determine numeral value of magic square corresponds to position of ASCII value, to provide additional security for the encryption system.

In 2015[6], Dawood, Rahma and Abdul Hossen presented a suggest algorithm that uses the Diffie-Hellman model to determine the dimensions of magic square block's physical structure which it determines the type of magic square. The properties of the magic cube were exploited in the processes of decryption / encryption & signing / verification. Magic cubes are based on 6 foldable magic squares that represent the towards of the magic cube.

In the year 2016 [7], Abdul-Mouniem, Abdul-Mohsien and Dawood presented cryptography algorithm based on Diffie-Hellman key exchange protocol , magic square and the Secure Hash Algorithm SHA-1 ,as the proposed algorithm relies on encryption/ decryption mainly on the magic constant and magic sum of the matrix, which is obtained after multiplying the matrix by a randomly selected constant value.

In the year 2016[8], researchers presented new approach to build a magic cube by using the folded magic square technique, where 6 magic squares are constructed in sequence or with a fixed different value between each pairs of numbers in the square mould, then each square is arranged in the correct order to preserve the properties of the magic cubic.

In 2018[9], Jabbar and Rahma presented cryptography approach Using the magic square 4×4 , linear equation system and finite field. So that the main dependence was on the magic sum.

In 2019[10], AL-Hashemy and Mahdi presented algorithm to encryption color images by dividing the images into small parts equivalent to the magnitude of the magic square that used and multiplying them by the magic square, and the XOR operation with a key was used in this proposed method.

In 2019, [11] Rahma and Jabbar presented protocol relies on magic square of size 3×3 , linear equation system and finite field, To maintain the integrity of the data when exchanged between two parties.

In 2020, [12] Chen presented new approach by Taking advantage of well-known magic square properties. Innovative way to search for prime numbers has been developed by engage the magic square and Goldbach's Conjecture.

In 2020, [13] Mohammed and Hasan presented new cryptography approach To prevent attackers from exploiting redundancy weaknesses in the cipher text, the magic square was used to provide more permutation and also to ensure the availability of an inverse matrix that could be used in the decryption process.

3. Advantages of previous technologies

Magic squares have several features where the sum of the numbers in any column, row, or diameter is equal, This resulting number is called magic constant or the magic sum. A square with the above properties is called a normal magic square, while a magic square is called a mini magic square, if the sum of any column or row in it is equal only (not including the diagonals)[2].

To take advantage of the characteristics of the magic square in the cryptography, the transition from the unspecified number group (from 1 to 25 in MS 5) was replaced in the finite numeric group $GF(P)$ where p is the prime number, and all the Galois field then will depending on (p) [14].

And the inability to benefit from applied it on devices that are currently available (Laptop, mobile, tablet,...) that depend on 8 bits We had to move to $GF(28)$ [14].

Therefore, assume that want to use a cryptography algorithm that works on 8-bit data every time, and want to perform the division operation with 8 bits, the representation of integers will be in the range of $0 - 255$, However, will note that 256 is not a prime number, and if the mathematical operations are performed in $Z_{256} \pmod{256}$ then this group of integers will not be field, Because the lowest prime number of 256 will be 251, so the group Z_{256} using the arithmetic mod is field, and the group from 252 - 255 is ineffective (discard), It gets us trouble so the polynomial numbers was used to get rid of this problem[14].

polynomial numbers which are expression consisting of variables and coefficients, all the operations that are executed are taken to module $m(x)$ so $m(x)$ is mod reduction of the irreducible polynomials which has the accuracy same properties as the prime number in the integer calculation and reducible polynomial means that can be analyzed into its factors of a smaller degree than it, while irreducible

polynomial, which cannot be analyzed into factors smaller than it, is divisible only by itself and by one [14].

A linear equation system, Gaussian elimination and finite field were used. Before applying the Gaussian elimination the equations must being arranged in such a ways that will be the diameter does not contain any zero and note that the numbers of equations must be equal to the number of unknowns that to obtain the results[9].

Cryptography represent the process of any converting process of the plaintext into hard to understand cipher text with a key. Encrypted text can only be recovered into plain text if the key is known, and it is secret but known only to parties that exchange information. In cryptography, a substitution cipher is work on fixed system which is units of plaintext are exchange with cipher text; it can be a single bit or a group of bits[14].

This paper will use a composite method among rules of MS and GF both its types.

4. The proposed cryptography technique:

By taking advantage of MS 5 property, where the key was used and placed in the positions that were agreed upon, then the message is placed in the remaining positions and then the sum is found, after that each sum is converted into an equation, so 12 equations were obtained (5 columns + 5 rows + 2 diagonal) as is As shown in Figure 1, then the process is repeated each time according to the message length used.

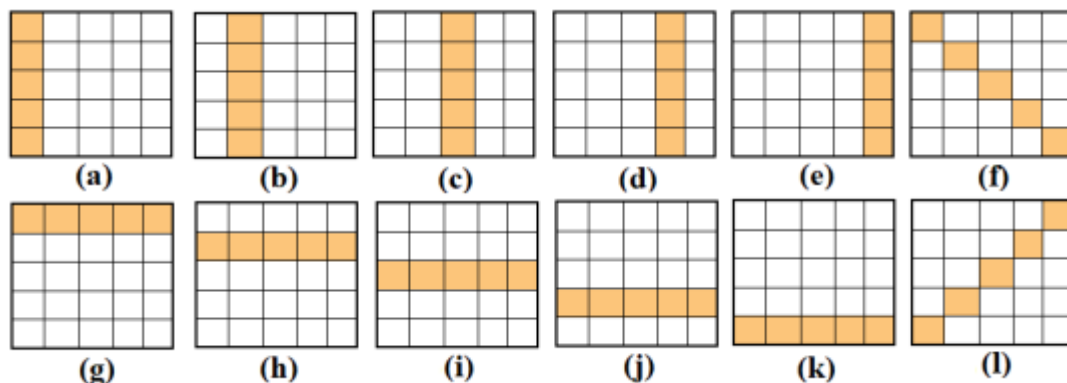


Figure 1. The positions in the Magic square 5×5 used as equations.

4.1 Key position determine by the user with length of message equal to 10:

When examining the 12 equations, it was noticed that they have dependability, so two equations (Figure 1 c&i) were cancelled, leaving 10 equations.

Therefore, for that will be 15 locations for keys, and the remaining parts will be 10 locations for messages (corresponding to 10 equations), and the result is 25 positions for Magic Square 5×5 .

Note that the proposed algorithm does not restrict the locations of the keys, but rather gives flexibility in choosing the key locations and the key value.

4.1.1 inserting the key and message It is as example that the following key's locations have been chosen as shown below:

	k	k		
k		k	k	
k	k		k	k
k	k			k
	k	k	k	

Figure 2. The key positions (assume) for length=10.

With the figure 2 and the remaining positions will contain the message, matrix S will be formed which consist from key and message as shown below:

S_{00}	S_{01}	S_{02}	S_{03}	S_{04}
S_{10}	S_{11}	S_{12}	S_{13}	S_{14}
S_{20}	S_{21}	S_{22}	S_{23}	S_{24}
S_{30}	S_{31}	S_{32}	S_{33}	S_{34}
S_{40}	S_{41}	S_{42}	S_{43}	S_{44}

Figure 3. The key (Shaded areas) and message (the remaining) in matrix S

Therefore, sums are obtained as shown by the equations below:

$$\begin{aligned}
 \text{Sum1} &= S00+ S01 + S02 + S03 + S04 \\
 \text{Sum2} &= S10+ S11 + S12 + S13 + S14 \\
 \text{Sum3} &= S30+ S31 + S32 + S33 + S34 \\
 \text{Sum4} &= S40+ S41 + S42 + S43 + S44 \\
 \text{Sum5} &= S00+ S10 +S20 + S30 + S40 \\
 \text{Sum6} &= S01+ S11 + S21 + S31 + S41 \\
 \text{Sum7} &= S03+ S13 + S23 + S33 + S43 \\
 \text{Sum8} &= S04+ S14 + S24 + S34 + S44 \\
 \text{Sum9} &= S00+ S11 + S22 + S33 + S44 \\
 \text{Sum10} &= S04+ S13 + S22 +S31+ S40
 \end{aligned} \tag{1}$$

So the sums resulted will be represented the encrypted text and then will be sent to the receiver.

The receiver will resolve the 10 equations with the 10 other unknowns and with 15 key vales as we explained in this paper (2-a - the 2nd paragraph).

And take into account before solving the equations make rearrange the equations, So the main diameter values don't equal to the zero.

The Solving the equations will represent the message.

These proposed works were developed to increasing the difficulty and strength of breaking it so the polynomial numbers were used instead of numbers in both the key, message and value.

4.1.2 the proposed algorithms. For that there is no stable or specific algorithm, this proposed approach was developed using MS 5 for decryption and encryption

Algorithm 1: The Proposed Symmetric Encryption algorithm based on MS5

Input: Plaintext message and key and key position.

Output: Cipher text.

Begin:

Step1 : placing the keys at the agreed positions in MS5.

Step2 : placing the message at the other remaining positions in MS5.

Step3 : After applying steps 1 and 2, we get the matrix A (See fig. 3) and will find the sums of each equations (See (1)) .

And these totals will represent the ciphertext and sent to the receiver.

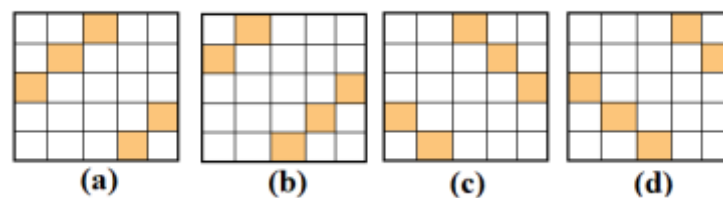
End.

Algorithm 2: The Proposed Symmetric Decryption algorithm based on MS5**Input:** Ciphertex and key and key position.**Output:** plaintext**Begin:****Step1:** placing the keys value in the chosen positions .**Step2:** The other positions will be remain Anonymous and their number is equal to the number of ciphertex = 10 , and ten equations will be formed, Depending on the matrix S (fig. 3) and (1).**Step3:** rearrange the result of equations so that the main diameter must not contain any value is equal to zero.**Step4:** Solve the 10 equations and relies on rules of the finite field to find the original text (the message), as shown in this paper (section 3 Paragraph 6).

The solve of the equation will represent the message.

End.**4.2 adding 4 new equations:**

Development to the work done to the length of message = 10 by adding four another extra equations, So the total sum of the equations will be = 14, The added equations can be illustrated in the following figure:

**Figure 4.** The equations added to the work.

4.2.1 inserting the key and message. It is assume the following key has been chosen :

	k	k	k	
k				k
		k		
k			k	k
k			k	

Figure 5. The key positions (assumed) for length=14.

The work will be similar as in the length of the message = 10 (section 3-1-1) and the difference will be in the added equations and the ten equations will be the same as previously used.

The added equations will be as shown below:

$$\begin{aligned}
 \text{Sum11} &= S02 + S11 + S20 + S34 + S43 \\
 \text{Sum12} &= S01 + S10 + S24 + S33 + S42 \\
 \text{Sum13} &= S02 + S13 + S24 + S30 + S41 \\
 \text{Sum14} &= S03 + S14 + S20 + S31 + S42
 \end{aligned}
 \tag{2}$$

4.2.2 the proposed algorithms for length = 14. The text of the algorithm is the same in the length of message equal to 10 (see section 3-1-2 in this paper).

The difference will be the number of equations will be 14 equations instead of 10 which equal the length of the message, on the other hand the length of the key is 11 locations to be the total 25 position (due to MS 5).

5. EVALUATION

Cryptography refers to the science that provides a secure connection and is associated by the process of converting the plaintext into ciphertext with the help of the key, and therefore without knowing the key, the third party cannot extract the message from the ciphertext, so it will be interesting to calculate the number of attempts to calculate the plain text by the third party which is called brute force attack.

Therefore, in this paper we used MS 5 with two different message lengths. It is considered an improvement to MS 3 and MS 4 by increasing the complexity and length of the key and thus it will be very difficult to reach the Plaintext.

In this section, we will calculate several parameters (complexity, encryption time, throughout, NIST tests and histogram analysis) for suggest MS 5 as shown below , Statistics were performed on the laptop Intel(R) Core(TM) i5, OS: Win10 ProX64, and by using C# in Microsoft Vis. Studio 2013.

5.1 Security complexity analysis

First, we'll calculate the key complexity represented by the brute force attack, as it will be in the first proposed algorithm P15 since P is the prime number used and 15 is the number of keys used in magic square. We will use a different prime number starting from 251 and incremental.

So also, for the 2nd suggested algorithm would be the complexity of the Brute force attack will be P11, since there are 11 keys in the magic square and we will also use different numbers of Prime Number starting from 251 onwards.

Secondly, we will calculate the complexity of the data for the first algorithm proposed, since there are 256 ASCII character in the data used and Message length will be = 10, then the complexity will be $= (256)^{10}$.

So also for the 2nd algorithm suggested the complexity of the data will be $(256)^{14}$, since the message length = 14 and 256 represent the character of ASCII.

So it will be the total complexity in the first and second algorithm equal to the complexity product of the data multiplied by the complexity product of the key for each one of them, as shown in (3) and (4) and the figures 6&7&8&9 will illustrate the relationship between the complexity of the two proposed algorithms with prime number =251 onwards and MS 3.

$$\text{Total complexity in the 1st algorithm} = (256)^{10} \times (251)^{15} \quad (3)$$

$$\text{Total complexity in the 2nd algorithm} = (256)^{14} \times (251)^{11} \quad (4)$$

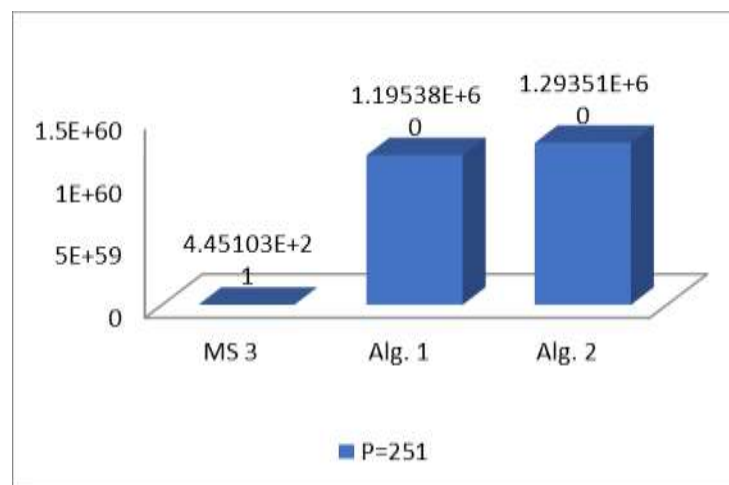


Figure 6. The Complexity in Algorithms 1 and 2 and MS 3 with prime=251.

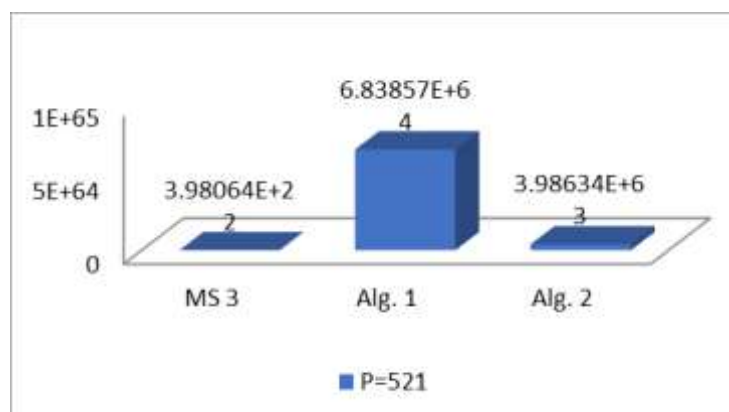


Figure 7. The Complexity in Algorithms 1 and 2 and MS 3 with prime=521.

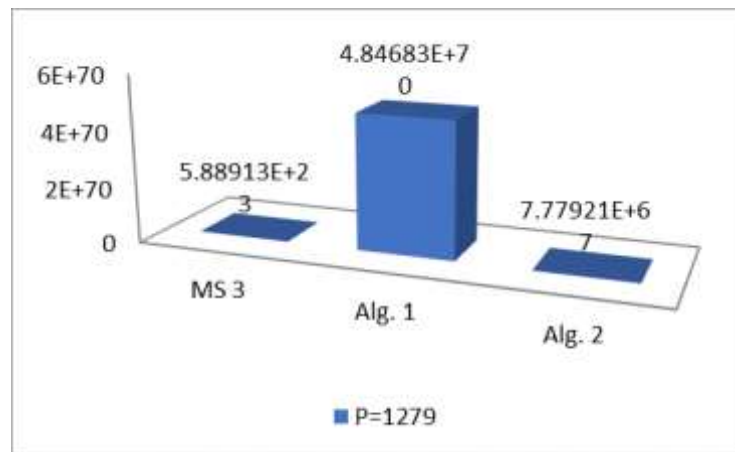


Figure 8. The Complexity in Algorithms 1 and 2 and MS 3 with prime= 1279.

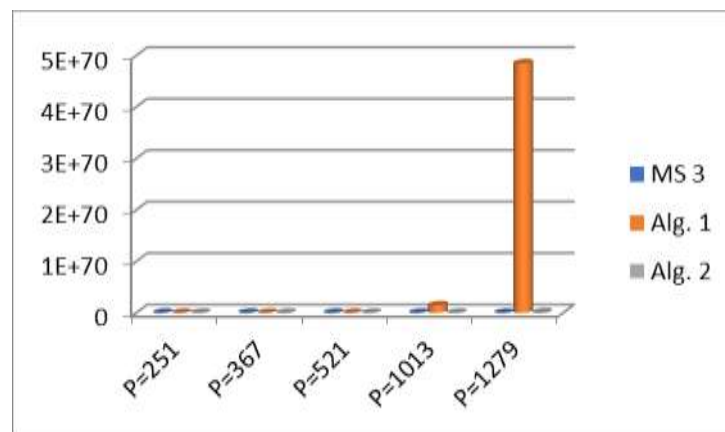


Figure 9. The Complexity in MS 3 and Algorithms 1 & 2 with different primes no.

5.2 Encryption and Decryption time

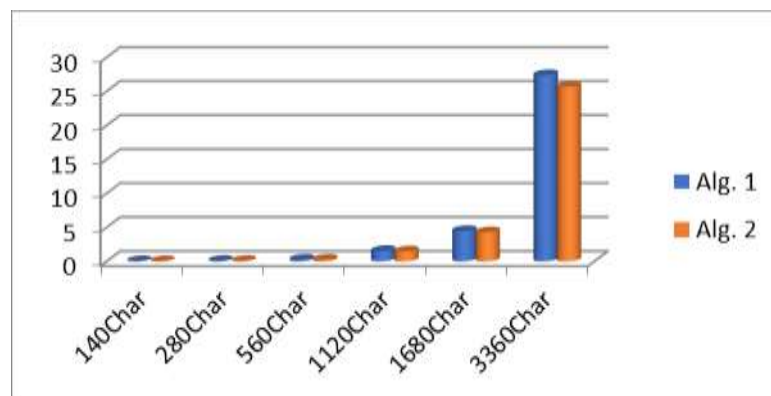
As another measuring for performance was calculated encryption and decryption time for each algorithm proposed where they were calculated for different sizes of message as shown in tables 1 & 2 and Figures 9 & 10 shows a comparison of the encryption and decryption times of the two algorithms for different sizes of messages.

Table 1. Execution times (in milliseconds) for Algorithm 1 to different Sizes of message.

No. of Char	Encryption (in unit m.s) Alg. 1	Decryption (in unit m.s) Alg. 1
140	00:00:00.0167418	00:00:00.0276373
280	00:00:00.0537937	00:00:00.0828052
560	00:00:00.2511645	00:00:00.3689929
1120	00:00:01.5166185	00:00:02.3846324
1680	00:00:04.4561050	00:00:06.4788159
3360	00:00:27.3825018	00:00:39.4362666

Table 2. Execution times (in milliseconds) for Algorithm 2 to different Sizes of message.

No. of Char	Encryption (in unit m.s) Alg. 2	Decryption (in unit m.s) Alg. 2
140	00:00:00.0160082	00:00:00.0264654
280	00:00:00.0522081	00:00:00.0814200
560	00:00:00.2436234	00:00:00.3725342
1120	00:00:01.4490555	00:00:02.2118289
1680	00:00:04.2634445	00:00:06.5029747
3360	00:00:25.6699058	00:00:39.4741034

**Figure 10.** Comparing between Alg.1 and Alg.2 in Encryption Time.

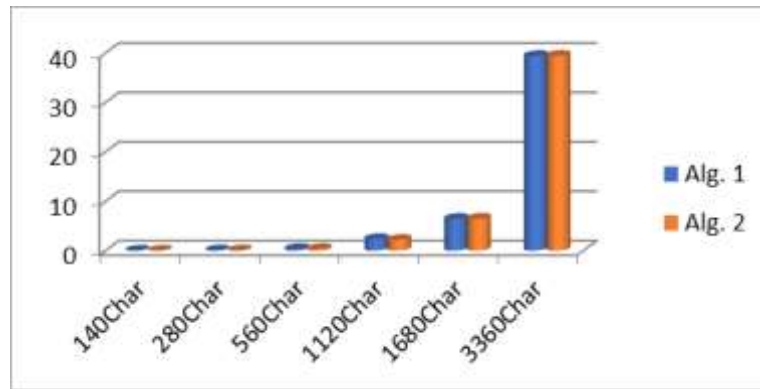


Figure 11. Comparing between Alg.1 and Alg.2 in Decryption Time.

5.3 Nist tests analysis

The output of plaintext encryption is preferred to be more random and difficult to predict, So it was measured NIST Test (National Institute of Standards and Technology) for each algorithm proposed as show in the tow tables 3 and 4.

Table 3. Some of NIST tests for Algorithm 1.

Statistical measures	Frequency	Cumulative sums	Runs	Longest of ones	runs Approximate entropy	Serial
Algorithm 1	0.900523	0.629222	0.382075	0.067103	0.602166	0.595074
Result	Success	Success	Success	Success	Success	Success

Table 4. Some of NIST tests for Algorithm 2.

Statistical measures	Frequency	Cumulative sums	Runs	Longest of ones	runs Approximate entropy	Serial
Algorithm 2	0.802586	0.338188	0.455502	0.018571	0.733584	0.568338
Result	Success	Success	Success	Success	Success	Success

And Table 5 illustrates some of ciphertext results in each of the two proposed algorithms, as shown below:

Table 5. Some of ciphertext results in Alg.1 & Alg.2.

Plain text	Ciphertext using Alg. 1	Ciphertext using Alg. 2
Computer Science	HúâynzîlQčkæ¾Æx	¼ŠÈ1ÁFöDİLñqLÔÉ
Programming in C#	ŊpLÝ÷đsđöč³ÚŊy²öljĚ	İRçû»űíŠkRâáĀNoáĀ

5.4 Discussion the results

It is evident by comparing the previous results in terms of complexity, implementation time and NIST test, that MS 5 is better in terms of randomness and complexity of the ciphertext than MS 3. Since the complexity ratio is very large in the proposed system in both methods compared to Magic Square 3, while the speed of implementation, the difference is not large, since the number of times the text is truncated each time will be more in MS 3 (because there are 9 locations) than in MS 5 (25 locations). This will be evident as the length of the message the user increases, on the other hand, there will be more equations in the MS 5 than in the MS 3, which takes longer, in any case increase in complexity is higher comparing to increase in execution time.

While for the comparison between the two proposed algorithms, it appears that the second algorithm is slightly superior to the first algorithm in terms of encryption time, but when comparing the complexity strength, we notice that the first algorithm is superior to the second algorithm, especially if the prime number grows, so the first algorithm is considered better than the second algorithm.

6. Conclusion :

The proposed algorithms were based on magic sum basically, having both the key and the message respectively placed in the magic square and that showed a good idea, as the use of a larger Magic Square leads to higher complexity.

Using higher prime number P in $GF(P)$ leads to higher complexity as well. In this block cipher algorithm, if the key is greater than the data then the algorithm is of higher complexity. $GF(2^8)$ was used instead of $GF(P)$ to be compatible with the hardware that depend with 8 Bits.

References

- [1] Sh. A. Salman , S. F. Ibraheem , A. k. Faraj and N. G. Al-Saidi , "Secure Authentication Protocol Based on the Generated Magic Square", Journal of Iraqi Al-Khwarizmi Society (JIKhS), Volume:2, pages: 14-21, 2018.
- [2] Gan Yee Siang, Fong Wan Heng, Nor Haniza Sarmin, " Properties and Solutions of Magic Squares", Menemui Matematik (Discovering Mathematics), Vol. 34, No. 1: 63– 76 ,(2012).
- [3] Israa N. Alkallak, " Using Magic Square of Order 3 to Solve Sudoku Grid Problem" , Ibn Al-Haitham Jour. for Pure & Appl. Sci. , Vol. 26 (1) ,(2017).
- [4] J. Wang and L. Liu, "A Novel Chaos-Based Image Encryption Using Magic Square Scrambling and Octree Diffusing", Mathematics , Chaos-Based Secure Communication and Cryptography ,vol 10 (3), 457 , 2022.
- [5] A. Dharini, R.M. Saranya Devi, and I. Chandrasekar, "Data Security for Cloud Computing Using RSA with Magic Square Algorithm", International Journal of Innovation and Scientific Research, ISSN 2351-8014 Vol. 11 No. 2 Nov. 2014, pp. 439-444 © 2014 Innovative Space of Scientific Research Journals <http://www.ijisr.issr-journals.org/>,(2014).
- [6] Omar A. Dawood , Abdul Monem S. Rahma and Abdul Mohsen J. Abdul Hossen , " New Variant of Public Key Based on Diffie-Hellman with Magic Cube of Six-Dimensions" , (IJCSIS), Vol. 13, No. 10, (October 2015).
- [7] Omar A. Dawood , Abdul Monem S. Rahma and Abdul Mohsen J. Abdul Hossen , "Public Key Cipher with Signature Based on Diffie-Hellman and the Magic Square Problem", Eng. & Tech. Journal, Vol.34, Part (B), No.1,(2016).
- [8] Omar A. Dawood , Abdul Monem S. Rahma and Abdul Mohsen J. Abdul Hossen, " Generalized Method for Constructing Magic Cube by Folded Magic Squares ", I.J. Intelligent Systems and Applications, 2016, 1, 1-8, Published Online in MECS (<http://www.mecs-press.org/>), DOI: 10.5815/ijisa.2016.01.01,(2016).
- [9] Doaa Ayad Jabbar and Abdul Monem S. Rahma , " Proposed Cryptography Protocol based on Magic Square, Linear Algebra System and Finite Field ", Jour of Adv Research in Dynamical & Control Systems, Vol. 10, No. 10, (2018).
- [10] Rageed Hussein AL-Hashemy and Sadiq A. Mehdi , " A New Algorithm Based on Magic Square and a Novel Chaotic System for Image Encryption ", J. Intell. Syst. 2020; 29(1): 1202–1215, <https://doi.org/10.1515/jisys-2018-0404> , Received October 3, 2018; previously published online (February 1, 2019).
- [11] Doaa Ayad Jabbar and Abdul Monem S. Rahma , " Development cryptography protocol based on Magic Square and Linear Algebra System ", Journal of AL-Qadisiyah for computer science and mathematics, Vol.11 No.1, ISSN (Print): 2074 – 0204, ISSN (Online): 2521 – 3504,(2019).
- [12] Hui-Shan Li, Chenglian Liu, and Sonia C-I Chen, " A Study of Authenticated Communication Based on Magic Square and Goldbach's Conjecture" , INTERNATIONAL JOURNAL OF CIRCUITS, SYSTEMS AND SIGNAL PROCESSING, DOI: 10.46300/9106.2020.14.28, Volume 14, (2020).
- [13] Sahab Dheyaa Mohammed and Taha Mohammed Hasan , " Cryptosystems using an improving hiding technique based on latin square and magic square ", Indonesian Journal of Electrical Engineering and Computer Science, Vol. 20, No. 1, pp. 510~520, ISSN: 2502-4752, DOI: 10.11591/ijeecs.v20.i1.pp510-520, (October 2020).
- [14] Stallings, William, "Cryptography and network security: principles and practice 6 Edition," Person Education Inc, (2014).

Estimation of the reliability using Lognormal Distribution

Rabab Abdulrida Saleh ⁽¹⁾

Iqbal Mahmood alwan ⁽²⁾

`rabab.saleh@coadec.uobaghdad.edu.iq`

`Iqbal.alwan@coadec.uobaghdad.edu.iq`

Department of Statistics, College of Administration and Economics, University of
Baghdad, Baghdad, Iraq

Estimation of the reliability using Lognormal Distribution

Rabab Abdulrida Saleh ⁽¹⁾

Iqbal Mahmood alwan ⁽²⁾

Iqbal.alwan@coadec.uobaghdad.edu.iq rabab.saleh@coadec.uobaghdad.edu.iq

Department of Statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq

Abstract

This study uses the Lognormal distribution to assess the reliability of deturbidity and the level of residual chlorine in the water of Al Rashidiya Water Center in Baghdad, Iraq. The maximum likelihood method was used for estimating distribution parameters in the MATLAB program, and the Kolmogorov-Smirnov test for Goodness of Fit was applied at a significant level of 0.05. It was concluded that the average turbidity was 9.389 and the average was 9, which is within the legal range, the average chlorine was 0.0296 and the average was 0.03 which is well below the permissible level. The study also discovered that the average effectiveness of deturbation was 71.764%, which is sufficient but should be improved. the reliability for deturbation was 100% and for the residual chlorine 0.96, suggesting that the Rashidiya Water Center was reliable during the research period. The researchers recommend conducting frequent inspections of water treatment facilities and plants in Iraq and comparing failure distributions to determine the optimal distribution of all water impurities.

Keywords: Reliability, Efficiency, Lognormal Distribution, Maximum likelihood Estimator, Water Treatment Center.

1. Introduction

All drinking water treatment stations in Iraq work by legal limitations. As a result, frequent evaluations must be done to research and analyze the quality of these stations' outputs. Reliability can be broadly defined for a water treatment plant as the probability that a plant will meet the maximum allowable limits of the treatment standards within a specified time [1] One of the most widely used methods for assessing reliability is using the Lognormal distribution. The Lognormal distribution is a general probability distribution that can be used in assessing reliability and in assessing risk. for a Tigris River water treatment center. The site's water treatment center reliability in terms of depollution is affected by several factors; Among the most important of these factors are the wide variations in water quantity and quality, such as the amount of rain falling, and the torrents that end in the Tigris River. In the Rashidiya Water Center, one of the most important factors that hurt the operation of the center and thus reduce reliability is rain, torrential rains and illegal discharge of sewage water.[2]

Al Rashidiya Water Center is located in the north of Baghdad, about 200 meters away from the Tigris River, with a design capacity of 2000 m³/h [3], as the project's work mechanism includes drawing raw water from the river and then entering it into Relatively large ponds to be mixed with alum, after that the water is distributed to the sedimentation ponds, and then the water is transferred to the filtration

ponds, the depth of each basin ranges (12-14)meters, a length of 4 meters, and a width of 5 meters, each basin consists of several layers of gravel and sand of varying thickness. Then, the water is pushed into ground tanks, in which the final chlorine([CL] _2) sterilization process takes place.[4]

2. Method

2.1 Reliability

The study was conducted during starting from 16/6/2021 to 15/6/2022, on daily (365) days data measured according to modern devices turbidity and chlorine CL_2), where turbidity is measured before and after the treatment process, and the study relied on turbidity data after treatment only. Chlorine (CL_2) is measured after the treatment process to know the remaining amount of it in the treated drinking water, as it is used in the sterilization process.

The acceptable limits for discharged water after treatment are as follows [5][6]:

Turbidity: 15 NTU

Residual chlorine CL_2): 3 mg/L

The reliability of the Rashidiya water center was determined using the Lognormal distribution, the Lognormal distribution can be used when the failure rate (the Exceedance Rate, in the case of water treatment plant) follows a monotonic trend. probability density function of The Lognormal distribution is: A positive random variable X is lognormally distributed(i,e) $X \sim \text{Lognormal}(\mu_x, \sigma_x^2)$, if the

natural Lognormal of X is normally distributed

with two parameters mean μ and variance σ^2 [7] :

$$\mu = \ln \left(\frac{\mu_x^2}{\sqrt{\mu_x^2 + \sigma_x^2}} \right) \quad \dots (1)$$

$$\sigma^2 = \ln \left(1 + \frac{\sigma_x^2}{\mu_x^2} \right) \quad \dots (2)$$

x : turbidity or (chlorine) data.

μ : location parameter.

σ : scale parameter.

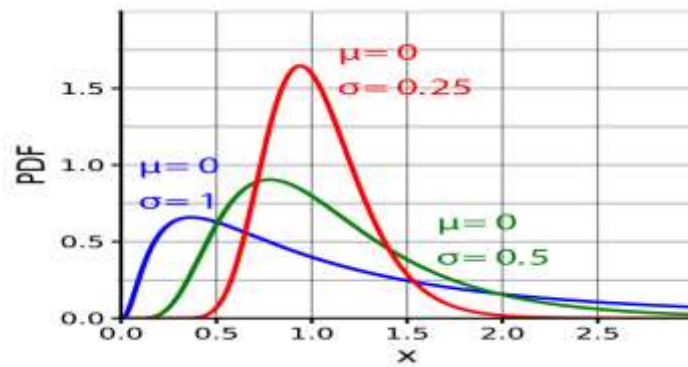


Figure 1: the probability density function of Lognormal distribution

The Cumulative distribution function is [8] ;

$$F_X(x) = \Phi\left(\frac{(\ln x) - \mu}{\sigma}\right) \quad \dots (3)$$

Where Φ is the cumulative distribution function of $N(0,1)$

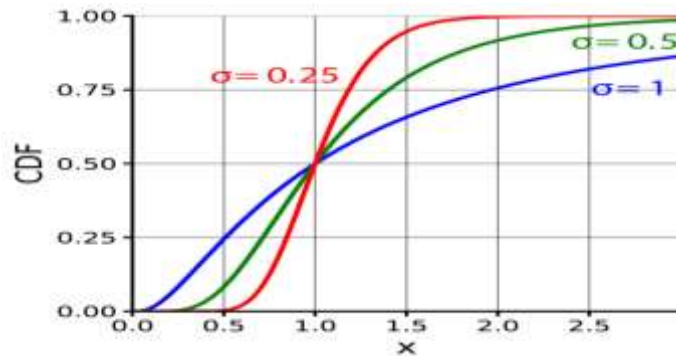


Figure 2: cumulative distribution function of the Lognormal distribution

The reliability function is given by [9]:

$$R(t) = 1 - \Phi\left(\frac{\text{Log}t - \mu}{\sigma}\right) \quad \dots (4)$$

We, however, want the Maximum Likelihood Estimator (MLE) μ and σ for the lognormal distribution, we can use the same procedure as for the normal distribution which is:

The log-likelihood function for a sample $\{x_1, \dots, x_n\}$ from a lognormal distribution is [7]:

$$LL = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (\ln x_i - \mu)^2 - \sum_{i=1}^n \ln x_i \quad \dots (5)$$

$$= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (\ln x_i)^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^n \ln x_i - \frac{n\mu^2}{\sigma^2} - \sum_{i=1}^n \ln x_i \quad \dots (6)$$

The log-likelihood function for a normal distribution is:

$$LL = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (\ln x_i - \mu)^2 \quad \dots (7)$$

Thus, the log-likelihood function for a sample $\{x_1, \dots, x_n\}$ from a lognormal distribution is equal to the log-likelihood function from $\{\ln x_1, \dots, \ln x_n\}$ minus the constant term $\sum \ln x_i$. Since the constant term doesn't affect which parameter values produce the maximum value of LL, we conclude that the maximum is achieved for the same values of μ and σ on the sample $\{\ln x_1, \dots, \ln x_n\}$ taken from a normal distribution, namely:

$$\mu = \frac{1}{n}\sum_{i=1}^n \ln x_i \quad \dots (8)$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^n (\ln x_i - \mu)^2 \quad \dots (9)$$

The MLE of reliability function is:

$$R(t)_{mle} = 1 - \Phi\left(\frac{\text{Log}t - \mu_{mle}}{\sigma_{mle}}\right) \quad \dots (10)$$

The mean, standard deviation and turbidity removal efficiency were calculated according to the following formula [1]:

$$eff = \left(1 - \frac{Turbidity_{out}}{Turbidity_{in}}\right) * 100 \quad \dots (11)$$

$Turbidity_{out}$: Turbidity after treatment.

$Turbidity_{in}$: Turbidity before treatment.

3. Results and discussion

Descriptive statistics were calculated, and it was found that the values of turbidity and chlorine were close to the acceptable standards according to the requirements [8] [8].

Table (1) Arithmetic Mean, St. deviation and CV for turbidity and residual chlorine.

Parameter	Unit	Mean	St. Deviation	CV
Turbidity	Unity	9.389	2.8520	0.3038
Chlorine	Mg/L	0.0296	0.0019	0.0625

To determine the reliability of turbidity and chlorine for Al-Rashidiya Water Center, using the log-normal distribution, and to ensure the goodness of fit, the Kolmogorov-Smirnov test was used at a significance level of $\alpha = 0.05$.

Table (2) the results of estimating of the lognormal distribution,

Parameter	Unit	Parameter Distribution		Kolmogorov-Smirnov	
		μ	σ	Test Value	P-Value
Turbidity	Unity	-3.5219	0.0736	0.23152	0
Chlorine	Mg/L	2.2232	0.1491	0.53294	0

Table (2) shows the estimators μ and σ^2 of the lognormal distribution, which were estimated using the Maximum likelihood Estimator (MLE). and the results of the Kolmogorov-Smirnov test for the experimental data.

Figure (3) shows that the reliability of meeting the specified turbidity threshold (15 units) was 0.96, with an exceedance probability of (0.04). This means that 362 times over the 365-day study period, a value equal to or less than 15 units was observed.

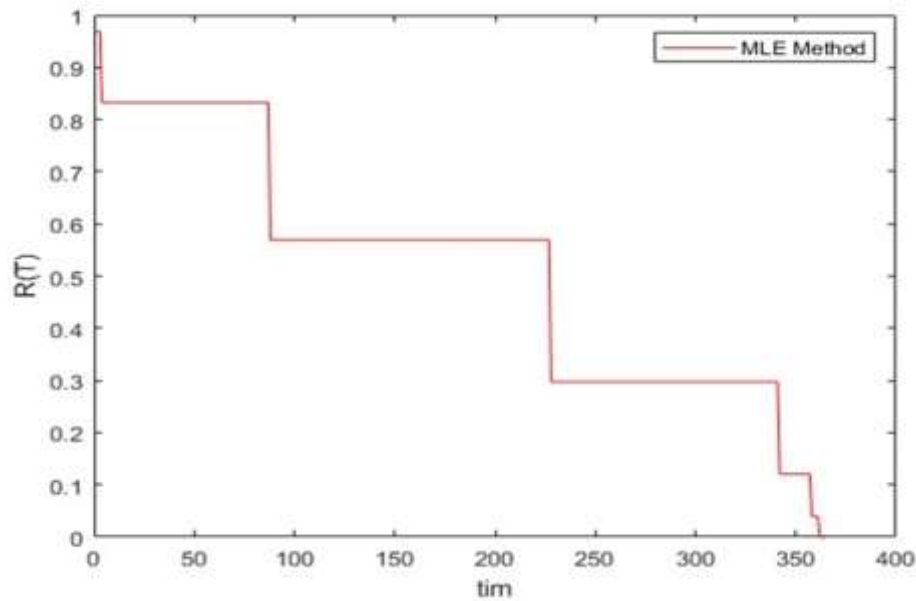


Figure 3: Turbidity reliability estimated using the Maximum likelihood Estimator (MLE) method.

Figure (3). depicts the dependability of estimations for turbidity levels using the Maximum Likelihood technique. The greatest reliability level was (0.96) at the smallest turbidity level, and the minimum reliability value was (0.04) at the maximum turbidity level, illustrating that reliability levels decrease as turbidity levels rise.

Graph of turbidity data in figure (4), we notice that the majority of the data falls within the permissible limit, and that some values fall outside the permissible limit, which is acceptable because the study period is relatively large, and the reason for this is due to rain and torrential rains that change the shape of the river water Tigris.

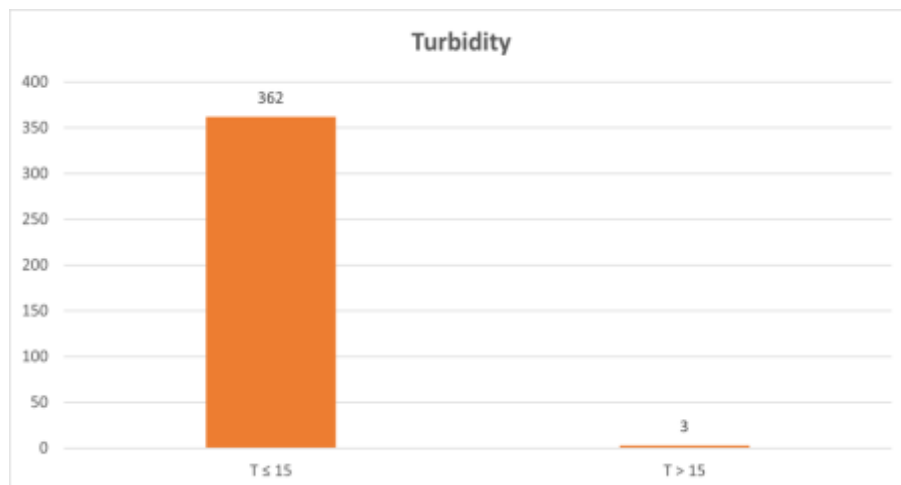


Figure (4) Graph of turbidity data.

Figure (4) shows a graph of the turbidity values and shows that (362) of the total values did not exceed the allowable treatment limit. While the permissible limit was exceeded (3) times only.

The reliability of lowering chlorine to the required standards (3 mg/L) was (100%), as indicated in Figure (5) with a chance of exceeding (0). This shows that no value of (3 mg/L) or above was reported over the 365-day research period.

Figure (5) depicts the Maximum Likelihood method reliability estimates; it's apparent from the figure that reliability amounts decrease as chlorine values increase, as the greatest reliability value was (1) at the smallest chlorine value and the smallest reliability value was (zero) at the greatest turbidity value.

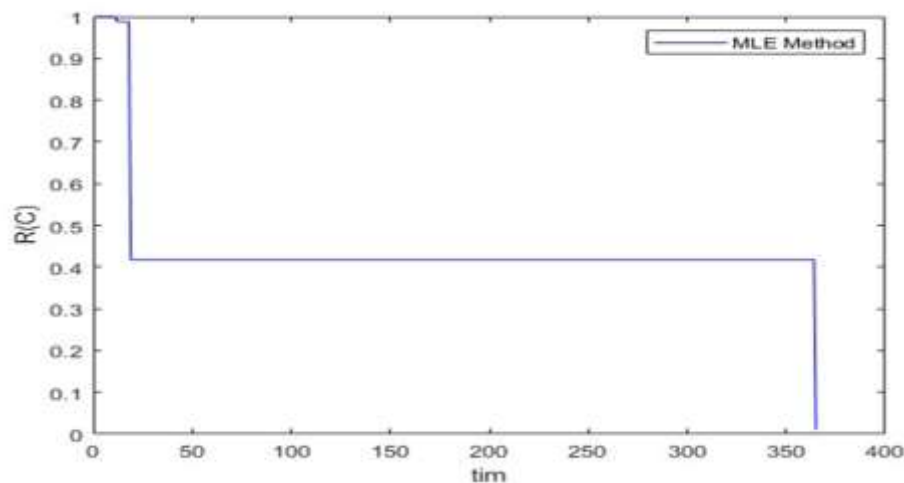


Figure 5: Chlorine reliability estimated using the Maximum likelihood Estimator (MLE) method.

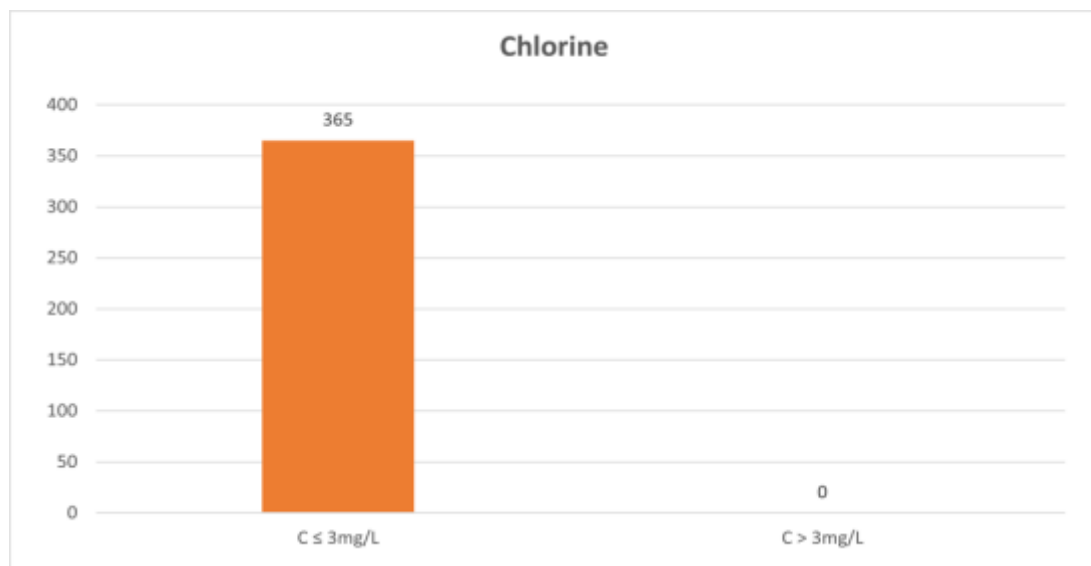


Figure (6) Chlorine data graph.

Because there were no cases of excess during the 365-day research interval, the chlorine measurements graph in Figure (6) shows that each value was less than the treatment-allowed limit.

4. Discussion

The study also discovered that the average turbidity removal effectiveness was 71.764%, which is adequate but has to be improved. The analysis revealed that the reliability value for turbidity removal was 100% and for remaining chlorine 0.96, suggesting that the Rashidiya water center was dependable during the research period. The report advises conducting frequent inspections of water treatment facilities and plants in Iraq and comparing failure distributions to determine the optimal distribution of all water impurities.

References

- [1] S. Oliveira and M. Sperling, "Reliability analysis of wastewater treatment plants," *Water Res*, vol. 42, no. 4–5, pp. 1182–1194, 2008.
- [2] P. Bugajski and E. Nowobilska-Majewska, "A Lognormal Analysis of the Reliability of a Wastewater Treatment Plant in Nowy Targ, Poland," 2019.
- [3] H. Algretawee, R. J. Mohammad, and M. H. Jassim, "Physical Evaluation for the Raw Water of Tigris River in Baghdad City," 2013. [Online]. Available: <https://www.researchgate.net/publication/329643062>.
- [4] M. B. F. and A.-N. B. A. B. Al-Obaidi, "Turbidity and its removal efficiency in the main water purification plants in Nineveh Governorate," *Mesopotamia Journal*, vol. 24, no. 3, pp. 39–53, 2013.
- [5] The Central Agency for Standardization and Quality Control, "Iraqi Standard Specifications for Drinking Water," 2001.
- [6] A. Y., Al-Safawi, M. A. Abdulhafedh, and M. K. AL-Taay. "Assessment of drinking water quality at the University of Mosul using the WQI Water Quality Index," *Journal /Scientific Studies (KUJSS)*, vol. 13, no. 2, pp. 185–198, 2018.
- [7] Genos, B. F. (2009) *Parameter estimation for the Lognormal distribution*
<https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=2927&context=etd>
- [8] M. Taboga, "Log-normal distribution", *Lectures on probability theory and mathematical statistics* (2021), Kindle Direct Publishing. Online appendix.
- [9] A. Katz, C. Williams, and J. Khim, Brilliant: Log-normal Distribution (2022), *retrieved on* 2022–02–06.

Estimation of Poisson Regression Model Using Robust Mallows Quasi-Maximum likelihood Method

A.Prof.Dr.Iqbal Mahmoud Alwan

Asaour Ammar

Ahmed The Mashhadani

Iqbal.alwan@coadec.uobaghdad.edu.iq

asawir.ammar2201m@coadec.uobaghdad.edu.iq

جامعة بغداد/كلية الادارة و الاقتصاد/قسم الاحصاء

Estimation of Poisson Regression Model Using Robust Mallows Quasi-Maximum likelihood Method

A.Prof.Dr.Iqbal Mahmoud Alwan

Asaour Ammar Ahmed The Mashhadani

asawir.ammar2201m@coadec.uobaghdad.edu.iqIqbal.alwan@coadec.uobaghdad.edu.iq

جامعة بغداد/كلية الادارة و الاقتصاد/قسم الاحصاء

Abstract:

In this research, the Poisson regression model is discussed in the case of a data set containing extreme values, and ways to overcome this problem were discussed using two robust methods (the weighted maximum likelihood estimation method and the robust mallows quasi-maximum likelihood estimation method) and a comparison between the two estimation methods to overcome the problem of extreme values and to know which is better. A sample size of (100) people was chosen and (6) tests were taken that diagnose Lupus disease or lupus erythematosus for the year (2022), as this real data was collected from private laboratories affiliated with the Medical City and the Iraqi Ministry of Health. It was shown through the results that the extreme values were treated through the two robust methods, and the best method was the robust mallows quasi-maximum likelihood estimation method.

Keywords : Generalized linear models , Poisson regression , Outliers , Robust regression methods , Robust mallows quasi-maximum likelihood (MQL) , Weighted maximum likelihood (WML) .

الدراسات السابقة :

❖ في عام (2011م) قام كل من (تارا و مهدي)^[1] باستخدام الانحدار الحصين لايجاد أنسب نموذج لتمثل بيانات الأنواء الجوية في مدينة اربيل خلال فترة (1998-2010) لتطبيق طريقة الانحدار الحصين في حالة وجود الشواذ، وتحديد مستوياته و التحري عنه، وكيفية التعامل معه في حالة وجوده في البيانات، ذلك بالاعتماد على إحصائية (DIFFTS,Leverage) و الطريقة المعروفة بطريقة بوكس ويسكر بلوت (Box and Whisker Plot) و ذلك للوصول الى ما تحاول الباحثين أن تثبته وهو ان استخدام الطرائق الحصينة افضل لتقدير معلمات النموذج الخطي سواء كانت البيانات تتوزع توزيعا طبيعيا ام لا في ظل وجود الشواذ في البيانات ام لا. ولهذا الغرض تمت دراسة الرطوبة النسبية والعوامل المؤثرة فيها . و تم جمع البيانات من دائرة الانواء الجوية في اربيل، وتتألف العينة من بيانات شهرية مسجلة خلال السنوات (1998-2010) و بحجم 12 مشاهدة لكل سنة. و تم استخدام الحقيبة البرمجية الجاهزة (Statgraphics, version 4(1999) و S-Plus, version 4,5(2000) و قد تبين من خلال النتائج استخدام الانحدار الحصين في حالة احتواء البيانات على شواذ .

❖ في عام (2020م) قام كل من (Omnia و Mohamed)^[18] دراسة مقارنة المقدرات الحصينة لنموذج انحدار بواسون مع القيم المتطرفة والتي تشمل مقدرات الامكان الاعظم الموزون WML ومقدرات شبه الامكان الاعظم الحصينة MQL ومقارنتها مع المقدرات الغير حصينة والتي تشمل مقدر الامكان الاعظم لنموذج بواسون مع القيم المتطرفة واجريت الدراسة محاكاة مونت كارلو باستخدام برنامج R وقد تبين من خلال نتائج المحاكاة أن مقدر الامكان الاعظم حساس بوجود القيم المتطرفة في حين مقدرين الامكان الاعظم الموزون و شبه الامكان الاعظم الحصينة اكثر فعالية بلأضافة الى ذلك أن مقدر الامكان الاعظم الموزون اكثر كفاءة من مقدر شبه الامكان الاعظم الحصينة .

❖ في لعام نفسة قام الباحث (مائل)^[2] مقارنة بين بعض الطرائق الحصينة والغير الحصينة مثل طريقة المربعات الصغرى أذ تم تحليل بيانات حقيقية أخذت من الجهاز المركزي للإحصاء المجموعة الإحصائية لعام 2017 والتي تخص الاحوال الطبيعية في العراق وتم استعمال برنامج الإحصائي Eviews 10 وقد تبين من خلال النتائج أن مقدرات الحصينة زودتنا بأفضل نموذج وفقاً للمقاييس الإحصائية .

❖ في عام (2023م) قامت الباحثة (سارة)^[3] مقارنة بين أنموذج انحدار بواسون وأنموذج ثنائي الحدين السالب للوصول الى أفضل نموذج أذ تم سحب عينة عشوائية بسيطة حجمها (100) حالة خاصة بمرض سرطان الثدي من وزارة الصحة والدوائر المتعلقة بها وكذلك مركز الابحاث السرطانية ودائرة مدينة الطب وقد تبين من خلال استعمال معيار متوسط مربعات الخطأ (MSE) ومعيار (AIC) أن أفضل نموذج هو نموذج بواسون .

المقدمة [19][15] :

أن التوزيع الاحتمالي الأكثر شيوعاً المستخدم في نمذجة بيانات العد هو توزيع بواسون Poisson Distribution . وأن وجود القيم المتطرفة Outliers في مجموعة بيانات تؤدي الى استنتاجات خاطئة خلال تحليل البيانات وتعتبر هذه احدى المشكلات الإحصائية التي تضعف التقديرات الغير حصينة مثل تقدير الامكان الاعظم و تقدير المربعات الصغرى الاعتيادية حيث تكون هذه التقديرات شديدة الحساسية تجاه القيم المتطرفة ويعد تحديدها ومعالجتها امر هاماً حيث من شأنها زيادة الخطأ ومن ثم عدم دقة التقديرات لذلك تم اقتراح طرائق التقدير الحصينة Robust Estimation Methods لنموذج انحدار بواسون للتغلب على هذه المشكلة.

وجود مشكلة في البيانات تحتوي قيم شاذة (متطرفة) في البيانات تتوزع توزيع بواسون .

هدف البحث :

يهدف هذا البحث لمعالجة مشكلة البحث حيث أن تم معالجة القيم الشاذة (المتطرفة) في البيانات التي تتوزع توزيع بواسون باستخدام طريقتين التقدير الحصينة والتي تشمل تقدير (WML, MQL) و المقارنة بين طريقتين التقدير الحصينة لمعرفة افضل طريقة لنموذج انحدار بواسون .

توزيع بواسون poisson distribution [22][12] :

عرف توزيع بواسون لدى عالم الرياضيات الفرنسي المشهور سيميون دينيس بواسون Simeon Denis Poisson هو من التوزيعات الاحتمالية المتقطعة يستخدم في الظواهر نادرة الحدث ويعرف كما في الصيغة الآتية :

$$f(Y_i = y/\mu) = \frac{e^{-\mu_i}}{y!} \mu_i^y \quad y = 0, 1, 2, \dots$$

$$i = 1, 2, 3, \dots$$

$$E(Y_i) = \mu_i$$

$$Var(Y_i) = \mu_i$$

نموذج انحدار بواسون Poisson Regression Model [3] :

يعد نموذج انحدار بواسون احد اهم النماذج الانحدار اللوغاريتمية الخطية وهوة الاداة التي يتم من خلالها نمذجة المتغير المعتمد عندما تكون قيم ذلك المتغير على شكل قيم قابلة للعد وكغيرة من سائر نماذج الانحدار قد يحتوي النموذج على المتغيرات مستقلة كثيرة مما يآثر سلباً على دقة النموذج وبساطته في تفسير النتائج و يفترض هذا النموذج أن المتغير المعتمد Y_i هو متغير استجابة يتبع توزيع بواسون وبمعلمه مقدارها μ_i كما تتبع الاخطاء العشوائية في النموذج توزيع بواسون بمعلمه قدرها μ_i ويعرف وفق دالة احتمالية :

$$Y_i = e^{(XB+U)}$$

ويمكن التعبير عنها ايضاً بصيغة المصفوفات

$$Y_i = \exp(XB + U)$$

Y_i : موجه المتغير التابع ذو درجة $(nx1)$ $i = 1, 2, 3, \dots$

X : مصفوفة المتغيرات المستقلة (التوضيحية) ذات الدرجة $(nx(p + 1))$

B : موجه المعلمات ذو الدرجة $((p + 1)x1)$

U : موجه الاخطاء العشوائية ذو درجة $(nx1)$

n : حجم العينة

P : عدد المتغيرات المستقلة (التوضيحية)

افتراضات نموذج انحدار بواسون [3] :

يعتمد النموذج انحدار بواسون على ثلاثة فروضات اساسية يتم الاعتماد عليها لبناء هذا النموذج :

الافتراض الاول :

أن الدالة الاحتمالية للمتغير المعتمد Y_i يتبع توزيع بواسون عندما تكون معلمة التوزيع μ_i

الافتراض الثاني :

أن معلمة التوزيع المتغير المعتمد Y_i تكون مساوي الى

$$=e^{X_i' B} \mu_i$$

الافتراض الثالث :

أن أزواج المتغيرين (Y_i, X_i) تكون بينهما استقلالية و أن الاعتماد على هذه الافتراضات الثلاثة مع خصائص توزيع بواسون سوف يكون الوسط الحسابي و التباين كما في الصيغة الآتية :

$$E(Y_i/X_i) = \mu_i = e^{X_i' B}$$

$$\text{Var}(Y_i/X_i) = \mu_i = e^{X_i' B}$$

$$E(Y_i/X_i) = \mu_i = e^{X_i' B} = \text{Var}(Y_i/X_i) = \mu_i = e^{X_i' B}$$

Poisson Regression Model and Maximum Likelihood Estimator نموذج انحدار بواسون ومقدر الامكان الاعظم [18] :

توزيع بواسون هو من التوزيعات الاحتمالية المتقطعة يستخدم في الظواهر نادرة الحدث ويعرف كما في الصيغة الآتية :

$$f(Y_i = y/\mu) = \frac{e^{-\mu_i}}{y!} \mu_i^y \quad y = 0, 1, 2, \dots \quad (1)$$

$$i = 1, 2, 3, \dots$$

في هذا التوزيع يجب ان يساوي المتوسط التباين وبالتالي لن يكون مناسب لبيانات بواسون بوجود القيم الشاذة [10] بالاعتماد على عينة (y_1, y_2, \dots, y_n) يمكننا كتابة النموذج بدلالة متوسط الاستجابة :

$$(2) E(Y_i) = \mu_i$$

$$= E(Y_i) + \varepsilon_i \quad (3) Y_i$$

$$= B_0 + B_1 X_{i1} + B_2 X_{i2} + \dots + B_p X_{ip} + \varepsilon_i \quad (4) Y_i$$

المعادلة (3) يمثل نموذج انحدار خطي عام وهو حالة خاصة من نماذج الخطية المعممة GLM ويكون الخطأ له توزيع معين وعادةً يكون $Normal \sim (0, \sigma^2)$ لكن في النماذج الخطية المعممة لم يعد هنالك خطأ فأن نفترض أن متغير الاستجابة Y_i له توزيع معين مثل توزيع بواسون وتوزيعات اخرى . تكون دالة الربط لتوزيع بواسون كما في الصيغة الآتية :

$$g(\mu_i) = \eta_i = \log(\mu_i) = B_0 + B_1 X_{i1} + B_2 X_{i2} + \dots + B_p X_{ip} = X_i' B$$

حيث أن X_i' هو متجه الصف المتغيرات المستقلة و B هو متجه عمود المعاملات

فإن الفكرة الأساسية لتقدير الامكان الاعظم جعل دالة الامكان للمتغير المعتمد في نهايتها العظمى اذا كان المتغير المعتمد يتبع توزيع بواسون كما في الصيغة (1)

فتكون دالة الامكان الاعظم كما في الصيغة الآتية :

$$L(y_1, y_2, \dots, y_n; \mu_i) = \frac{e^{(-\sum_{i=1}^n \mu_i)} \mu_i^{\sum_{i=1}^n y_i}}{\pi_{i=1}^n y_i!} \quad (5)$$

وبأخذ اللوغاريتم الطبيعي لدالة الامكان الاعظم لبواسون

$$L(y_1, y_2, \dots, y_n, \mu_i) = -\sum_{i=1}^n \mu_i + \sum_{i=1}^n y_i (\log(\mu_i)) - \log(\pi_{i=1}^n y_i!) \quad (6)$$

باستخدام الافتراض الثاني من الفروض الاساسية لنموذج انحدار بواسون

$$\mu_i = e^{X_i' B}$$

يتم التعويض في المعادلة (6)

$$\begin{aligned} L(y_1, y_2, \dots, y_n, \mu_i) &= -\sum_{i=1}^n e^{X_i' B} + \sum_{i=1}^n y_i (\log(e^{X_i' B})) - \log(\pi_{i=1}^n y_i!) \\ &= \sum_{i=1}^n (y_i X_i' B - e^{X_i' B} - \log y_i!) \end{aligned} \quad (7)$$

بتعظيم الاحتمالية ناخذ المشتقة

$$\frac{\partial \log L(B)}{\partial B} = \sum_{i=1}^n (y_i X_i' - X_i' e^{X_i' B} - 0)$$

باستخراج عامل مشترك X_i'

$$\sum_{i=1}^n X_i' (y_i - e^{X_i' B})$$

نساوي المعادلة بالصفر

$$\sum_{i=1}^n X_i' (y_i - e^{X_i' B}) = 0 \quad (8)$$

وبتالي فإن دالة الامكان الاعظم تؤدي الى المعادلة غير خطية وذلك سيتم استخدام المربعات الصغرى المعاد وزنها بشكل متكرر^[14] للحصول على تقديرات الامكان الاعظم كما في الصيغة الآتية :

$$\hat{B}_{ML}^{k+1} = (X' \hat{W}^k X)^{-1} (X' \hat{W}^k \hat{Z}^k) \quad (9)$$

\hat{W}^k : مقدر الوزن لـ μ_i

\hat{Z}^k : مقدر المتغير التابع المعدل

القيم الشاذة (المتطرفة) في البيانات Outlier in The Data^[2] :

هي الملاحظة التي تكون غير متسقة من حيث القيمة مع بقية المشاهدات وهي بذلك تبتعد عن مركز البيانات أن وجود القيم الشاذة في بيانات نموذج الانحدار يمكن أن تكون في المتغير المعتمد (متغير الاستجابة) ويطلق عليها (Outlier) ، او وجود القيم الشاذة في المتغيرات التوضيحية وعندئذ فإنها تسمى بنقاط الرفع (Leverage points) وهي على نوعين اما نقاط رفع جيدة والتي لا تؤثر على تقدير معادلة الانحدار او غير جيدة تؤثر على معالم النموذج ، كما أن وجود القيم الشاذة قد تكون في المتغير المعتمد وكذلك وجودها في المتغيرات التوضيحية معاً مما يفاقم من مشكلة التقدير ، فطرائق التقدير لنموذج الانحدار ومنها طريقة المربعات الصغرى تكون غير كفوءة حيث أن اخطاء النموذج لا تتبع التوزيع الطبيعي وهذا الأمر يتطلب استخدام طرائق الانحدار الحصين لمعالجة هذه المشاكل والحصول على مقدرات الكفاءة .

مفهوم طرائق الانحدار الحصينة The Concept of Robust Regression Methods^[1] :

تعد طرائق الانحدار الحصينة هي طرائق بديلة لطرائق الانحدار التقليدية مثل طريقة المربعات الصغرى الاعتيادية وطريقة الامكان الاعظم تستخدم طرائق الانحدار الحصينة في حالة وجود قيم متطرفة (شاذة) في البيانات حيث أن تقدير معالم نموذج الانحدار الاعتيادي في ظل وجود هذه الشواذ تكون غير كفوءة لأن تحدث في حالة عدم تطابق بين البيانات موضوع الدراسة والفروض الاساسية الواجب توفرها في النموذج وبذلك تفقد الطرائق التقليدية خصائصها الجيدة لتقدير معالم النموذج المدروس لذا تم ايجاد الطرائق الاحصائية بديلة تكون مقاومة لمعالجة هذه المشكلة وهي ما تسمى بلطرائق التقدير الحصينة . فإن هذه التقديرات ترتبط بتقديرات شبه الامكان^{[4][14][16][17]} Quasi Likelihood حيث أن تقدير شبه الامكان هو حل لنظام معادلات التقدير كما في الصيغة الآتية :

$$U(B) = \sum_{i=1}^n \frac{\partial}{\partial B} Q(\mu_i, y_i) = \sum_{i=1}^n \frac{\partial \mu_i (y_i - \mu_i)}{\partial B \text{Var}(\mu_i)} \quad (10)$$

حيث أن

$$\text{Var}(\mu_i) = \text{Var}(Y_i)$$

$$\frac{\partial \mu_i}{\partial B} = \left(\frac{\partial \mu_i}{\partial B_1} \frac{\partial \mu_i}{\partial B_2} \dots \frac{\partial \mu_i}{\partial B_p} \right) = \sum_{i=1}^n X_{ij} \frac{\partial \mu_i}{\partial \eta_i}$$

$$= \sum_{i=1}^n \frac{\partial \mu_i}{\partial \eta_i} \frac{(y_i - \mu_i)}{\text{Var}(Y_i)} X_{ij} \quad (11)$$

أولاً: تقدير الامكان الاعظم الموزون (WML-Estimation) [11][13][18]:

قدم Hosseinian and Morgenthaler [13] عام (2011م) احدى النماذج الحصينة هي طريقة التقدير الامكان الاعظم الموزون يشير WML الى (Weighted Maximum Likelihood) التي يمكن استخدامها لتقدير معلومات نموذج انحدار بواسون بلاعتماد على طريقة الامكان الاعظم المرجح بأوزان فإن المعادلة الناتجة هي الامكان الاعظم المرجح بأوزان تعتمد على μ_i و ثابتين C_1, C_2 .

يمكن الحصول على تقدير (Weighted Maximum Likelihood) باستخدام معادلة (11) ومن خلال التقليل باستخدام الأوزان الذي يقلل من تأثير المشاهدات التي تحتوي على قيم متطرفة حيث أن الوزن $W(\mu_i)$ يضمن المشاهدات ذات التأثير المرتفع تحصل على وزن اقل من المشاهدات ذات التأثير المنخفض كما في الصيغة الآتية:

$$= \sum_{i=1}^n \frac{\partial \mu_i}{\partial \eta_i} W(\mu_i) \frac{(y_i - \mu_i)}{\text{Var}(\mu_i)} X_{ij} \quad (12)$$

حيث أن

$$\text{Var}(\mu_i) = \mu_i$$

$$= \sum_{i=1}^n \frac{\partial \mu_i}{\partial \eta_i} W(\mu_i) \frac{(y_i - \mu_i)}{\mu_i} X_{ij}$$

$$= \sum_{i=1}^n \frac{\frac{\partial \mu_i}{\partial \eta_i}}{\mu_i} W(\mu_i) (y_i - \mu_i) X_{ij} \quad (13)$$

حيث أن

$$W(\mu_i) = \begin{cases} 1 & \frac{v}{C_1} < \mu_i < C_1 v \\ \frac{C_1 \mu_i}{v} & \mu_i < \frac{v}{C_1} \\ \frac{C_2 v - \mu_i}{v} & C_1 v < \mu_i < C_2 v \\ 0 & \text{otherwies} \end{cases}$$

v : الوسيط لـ μ_i

: C_1 2

: C_2 3

لتقدير معلمة الانحدار استخدم [13] Hosseinian and Morgenthaler طريقة نيوتن رافسون [14][17] كما في الصيغة الآتية :

$$\hat{B}_{ML}^{k+1} = (X' \hat{W}^k X)^{-1} (X' \hat{W}^k \hat{Z}^k)$$

$$\hat{B}_{ML}^{k+1} = B^k - \left(S'(B^k) \right)^{-1} S(B^k) \quad (14)$$

حيث أن

$$B^k = (X' W_i X)^{-1} (X' W_i \eta_i) \quad (15)$$

$$S(B^k) = \sum_{i=1}^n \frac{\partial \eta_i}{\partial \mu_i} (y_i - \mu_i) W_i X_{ij} \quad (16)$$

$$S'(B^k) = -X' W_i X \quad (17)$$

$$\hat{Z}^k = \left(\eta_i + \frac{\partial \eta_i}{\partial \mu_i} (y_i - \mu_i) \right) = \left(\hat{\eta}_i^k + (y_i - \hat{\mu}_i^k) \frac{\partial \hat{\eta}_i^k}{\partial \hat{\mu}_i^k} \right)$$

$$W_i = \hat{W}^k$$

ثانياً: تقدير شبه الامكان الاعظم الحصينة (MQL-Estimation) [17][18] :

اقترح Cantoni and Ronchetti [7] عام (2001م) تقدير حصين يعتبر حالة عامة من تقدير M ويشير MQLE الى (Mallows Quasi Likelihood Estimation) تقدير شبه الامكان الاعظم الحصينة حيث أن يتم تحديد تأثير القيم المتطرفة على المتغيرين X_i و Y_i بشكل منفصل و يستخدم هذا التقدير لتقليل من تأثير القيم المتطرفة كما في الصيغة الآتية :

نستخدم المعادلة كما في الصيغة (10)

$$\sum_{i=1}^n \frac{\partial \mu_i}{\partial B} \frac{(y_i - \mu_i)}{Var(\mu_i)}$$

حيث أن

$$\frac{\partial \mu_i}{\partial B} = \mu'_i$$

$$= \sum_{i=1}^n \mu'_i \frac{(y_i - \mu_i)}{Var(\mu_i)}$$

$$Var(\mu_i) = Var(Y_i) = \sigma^2$$

$$= \sum_{i=1}^n \mu'_i \frac{(y_i - \mu_i)}{\sigma^2}$$

$$= \sum_{i=1}^n \mu'_i \frac{(y_i - \mu_i)}{\sigma \times \sigma}$$

حيث أن

$$\sigma = \sqrt{\text{Var}(\mu_i)} = (\text{Var}(\mu_i))^{\frac{1}{2}}$$

$$= \sum_{i=1}^n \mu'_i \frac{(y_i - \mu_i)}{(\text{Var}(\mu_i))^{\frac{1}{2}}} (\text{Var}(\mu_i))^{-\frac{1}{2}} \quad (18)$$

حيث أن

$$r_i = \frac{(y_i - \mu_i)}{(\text{Var}(\mu_i))^{\frac{1}{2}}}$$

$$r_i = \Psi_c(r_i)$$

$$\Psi_c(r_i) = \begin{cases} r_i & , |r_i| \leq c \\ c \sin(r_i) & , |r_i| > c \end{cases}$$

حيث أن $\Psi_c(\cdot)$ دالة هايبرتم استخدام الثابت $C = 1.345$ للحصول على كفاءة نسبية بنسبة 95%

$$= \sum_{i=1}^n \mu'_i \Psi_c(r_i) (\text{Var}(\mu_i))^{-\frac{1}{2}}$$

حيث أن

$$v(y_i, \mu_i) = \Psi_c(r_i) (\text{Var}(\mu_i))^{-\frac{1}{2}} \quad (19)$$

$$= \sum_{i=1}^n \mu'_i v(y_i, \mu_i)$$

نستخدم الوزن $w(x_i)$ لتقليل من وزن نقاط المتغير المستقل المرتفعة حيث أن تم حساب $w(x_i)$ من قيم مصفوفة القبة لأن قيم مصفوفة القبة تتراوح من صفر الى الواحد

$$w(x_i) = \sqrt{1 - h_i}$$

h_i : هو العنصر القطري لمصفوفة القبة

$$H = X(X'X)^{-1}X'$$

Huber quasi-

عندما يكون الوزن يساوي واحد لكل مشاهدة نحصل على مايسمى بمقدر احتمالية هاير

likelihood estimator

$$w(x_i) = 1 \text{ for all } i$$

حيث أن $w(\cdot)$ و $\Psi_c(\cdot)$ تم اقتراحه كدالتين للحصول على مقدر حصين

$$= \sum_{i=1}^n \mu'_i w(x_i) v(y_i, \mu_i) - \alpha(B) \quad (20)$$

حيث أن $\alpha(B)$ هو Fisher-Consistency معامل التصحيح تم طرحه من الصيغة (20) للحصول على مقدرات غير متحيزة

$$a(\beta) = \frac{1}{n} \sum_{i=1}^n E[v(y_i, \mu_i)] w(x_i) \mu'_i$$

لتقدير معلمة الانحدار تم استخدام طريقة المربعات الصغرى المرجحة التكرارية [14][16][17] كما في الصيغة الآتية :

$$(X' \hat{W}^k \hat{Z}^k) \quad (21) \hat{B}_{ML}^{k+1} = (X' \hat{W}^k X)^{-1}$$

حيث أن

$$\hat{W}_i^k = \text{diag} \left(\frac{(\hat{\mu}_1^k)^2}{\phi \hat{\mu}_1^k} \dots \frac{(\hat{\mu}_n^k)^2}{\phi \hat{\mu}_n^k} \right)$$

$$\hat{Z}_i^k = \hat{\eta}_i^k + (y_i - \hat{\mu}_i^k) \frac{\partial \hat{\eta}_i^k}{\partial \hat{\mu}_i^k}$$

$$\hat{\mu}_i^k = g^{-1}(\hat{\eta}_i^k)$$

$$\hat{\eta}_i^k = x^T B^{k+1} = X B^{k+1}$$

البواقي للكشف عن القيم المتطرفة في انحدار بواسون (Residuals for detecting outliers in Poisson regression)

[21] : يعد فحص البواقي اسلوباً اساسياً لتحديد الاختلافات الاجمالية بين البيانات والنموذج مما يساعد على اكتشاف القيم المتطرفة في البيانات هناك انواع من البواقي المختلفة مثل بواقي الخام وبواقي بيرسون وبواقي بيرسون القياسية وبواقي الانحراف وبواقي الانحراف المعيارية للكشف عن القيم المتطرفة وايضاً يمكن استخدام هذه البواقي لقياس جودة النموذج. في هذا البحث تم استخدام بواقي بيرسون Parson Residual وبواقي الانحراف Deviance Residual للكشف عن القيم المتطرفة الموجودة في البيانات .

معيار المقارنة comparison standard^[18] :

بعد ان قدرنا معالم النموذج لكل طريقة من طرائق التقدير الحصينة نقوم باختيار افضل طريقة من بين هذه الطرائق من خلال معيار المقارنة متوسط مربعات الخطأ (MSE) Mean Square Error :

كما في الصيغة الآتية :

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{B} - B)^2 \quad (22)$$

\hat{B} : هو متجة القيم المقدرة

B : هو متجة المعاملات الحقيقية

n : حجم العينة

الجانب التطبيقي :

تم اختيار عينة حجمها (100) شخص وأخذ (6) تحاليل التي تشخص الإصابة بمرض لوباس او مرض الذئبة الحمراء لسنة (2022م) حيث أن تم تجميع تلك البيانات الحقيقة من مختبرات اهلية تابعة لمدينة الطب و وزارة الصحة العراقية سيتم تعريف كل متغير من متغيرات التي تم العمل عليها في الأنموذج .

B_0 : يمثل الحد الثابت

$B_1, B_2, B_3, B_4, B_5, B_6, B_7, B_8$: يمثل الميل الحدي

Y : يمثل اعداد المصابين بمرض الذئبة الحمراء وهو عبارة عن تحليل يؤخذ من بلازما الدم (Anti-Nuclear Antibody test) واختصار التحليل (ANA) .

X_1 : يمثل تحليل الاجسام المضادة للحمض النووي المضاعف (Anti-Double Stranded Antibody test)

واختصار التحليل (DNA) .

X_2 : يمثل تحليل الاجسام المضادة سميث (Anti-Smith Antibody) واختصار التحليل (SM) .

X_3 : يمثل تحليل الاجسام المضادة للفوسفوليبيد (Anti-Phospholipid Antibody Igm) واختصار التحليل (APL) .

X_4 : يمثل تحليل الاجسام المضادة للكارديوليبين (Anti-Cardiolipin Antibody IgM) واختصار التحليل (ACL) .

X_5 : يمثل تحليل مانع التخثر الذئبي (Lupus Anti-coagulant test) واختصار التحليل (LA) .

المتنبئون : ثوابت (Constant), الجنس (GENDER), العمر (AGE), DNA, SM, APL, ACL, LA

المتغير التابع : ANA

بواقي بيرسون Prarson Residual :

$$p_i = \frac{(y_i - \hat{\mu}_i)}{(Var(\hat{\mu}_i))^{\frac{1}{2}}} \quad (23)$$

$$Var(\hat{\mu}_i) = \hat{\mu}_i$$

بواقي الانحراف Deviance Residual :

$$d_i = \sin(r_i) \sqrt{2 \left[y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - r_i \right]} \quad (24)$$

بأستخدام برنامج R (إصدار 4.4.1) حيث كانت النتائج كما في الآتي :

بواقي بيرسون Prarson Residual	بواقي الانحراف Deviance Residual
900.658	1346.262

الجدول الأول

من خلال قيم البواقي والتي تعد قيم كبيرة مما يشير الى وجود قيم متطرفة في انحدار بواسون ولمعالجة هذه المشكلة تم استخدام طريقتي التقدير الحصينة التي تشمل (طريقة تقدير الامكان الاعظم الموزون و طريقة تقدير شبه الامكان الاعظم الحصينة) .

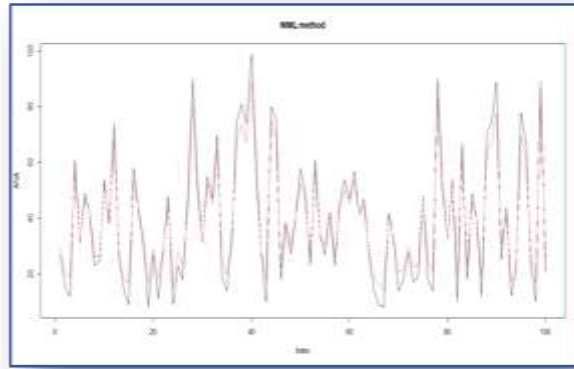
أولاً: تقدير الامكان الاعظم الموزون (WML-Estimation) :

MODEL	B	Std. Error	z	Sig
intercept	3.63277	0.01215	298.99	2e-16
LA(x_5)	0.06292	0.01302	4.83	1.40E-06
ACL(x_4)	-0.02255	0.01251	-1.8	0.0714
APL(x_3)	-0.00528	0.01464	-0.36	0.71824
SM(x_2)	-0.08702	0.01242	-7.01	2.50E-12
DNA(x_1)	-0.03191	0.01253	-2.55	0.01088
GEN	-0.04748	0.01327	-3.58	0.00034
AGE	-0.03275	0.01409	-2.32	0.02014

الجدول الثاني

يتضح لنا من الجدول اعلاه ان كلا من (AGE,DNA,SM ,LA,GEN) هي متغيرات معنوية كون قيمة p-value الخاصة بها اقل من 0.05 مؤثرة في حدوث المرض اذ نلاحظ ان LA يؤثر بشكل طردي على حدوث المرض اذ ان زيادته بمقدار وحدة واحدة يؤدي الى زيادة المرض بمقدار (0.06292) في حين كان تأثير كلا من (DNA و SM و AGE و GEN) لها تأثير عكسي اذ ان زيادتهما بمقدار وحدة واحدة يؤدي الى انخفاض المرض بمقدار (-0.08702) و(-0.03191) و (-0.04748) و (-0.03275) تواليا وفي ادناه الرسم التوضيحي للقيم الحقيقية لمتغير الاستجابة والقيم التقديرية . حيث أن نموذج الامكان الاعظم الموزون كما في الصيغة الآتية :

$$y = 3.63277 + 0.06292 \text{ LA} - 0.02255 \text{ ACL} - 0.00528 \text{ APL}$$



(1)

ثانياً: تقدير شبه الامكان الاعظم الحصينة (MQL-Estimation) :

MODEL	B	Std. Error	z	Sig
intercept	3.63277	0.01215	298.99	2e-16
LA(x_5)	0.06292	0.01302	4.83	1.40E-06
ACL(x_4)	-0.02255	0.01251	-1.8	0.0714
APL(x_3)	-0.00528	0.01464	-0.36	0.71824
SM(x_2)	-0.08702	0.01242	-7.01	2.50E-12
DNA(x_1)	-0.03191	0.01253	-2.55	0.01088
GEN	-0.04748	0.01327	-3.58	0.00034
AGE	-0.03275	0.01409	-2.32	0.02014

الجدول الثالث

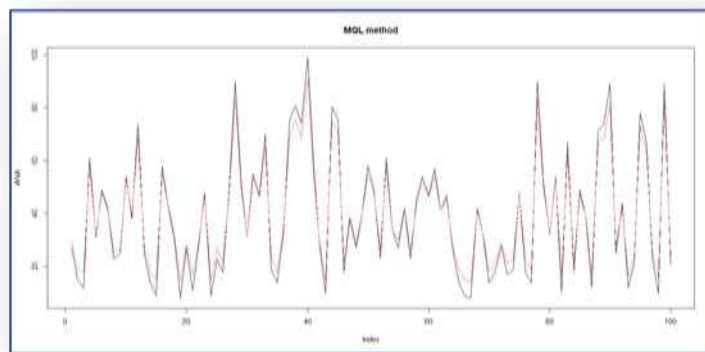
يتضح لنا من الجدول اعلاه ان كلا من (AGE,DNA,SM ,LA ,GEN) هي متغيرات معنوية كون قيمة p-value الخاصة بها اقل من 0.05 مؤثرة في حدوث المرض اذ نلاحظ ان LA يؤثر بشكل طردي على حدوث المرض اذ ان زيادته بمقدار وحدة واحدة يؤدي الى زيادة المرض بمقدار (0.06292) في حين كان تأثير كلا من (SM و DNA و AGE و GEN) لها تأثير عكسي اذ ان زيادتهما بمقدار وحدة واحدة يؤدي الى انخفاض المرض بمقدار (-0.08702) و (-0.03191) و (-0.04748) و (-0.03275) تواليا

وفي ادناه الرسم التوضيحي للقيم الحقيقية لمتغير الاستجابة والقيم التقديرية . حيث أن نموذج شبه الامكان الاعظم الحصينة كما في الصيغة الآتية :

$$y = 3.63277 + 0.06292 \text{ LA} - 0.02255 \text{ ACL} - 0.00528 \text{ APL}$$

(2) معيار المقارنة متوسط مربعات الخطأ (MSE) :

الطريقة	متوسط مربعات الخطأ MSE
MQL	13.46
WML	23.58



الجدول الرابع

قد تبين من خلال النتائج أن أقل قيمة متوسط مربعات الخطأ لطريقة التقدير شبه الامكان الاعظم الحصينة .

النتائج :

- ❖ تعتبر التقديرات الحصينة أكثر كفاءة من التقديرات التقليدية في حالة وجود قيم متطرفة في البيانات والعكس صحيح .
- ❖ أن أفضل طريقة تقدير هي شبه الامكان الاعظم الحصينة من خلال معيار المقارنة متوسط مربعات الخطأ فقد حقق أقل متوسط مربعات خطأ .

التوصيات :

- ❖ يوصى بتطبيق توزيع و انموذج بواسون على بيانات العد لجودته .
- ❖ يوصى باستخدام طرائق التقدير الحصينة في حالة احتواء البيانات على قيم متطرفة .
- ❖ يوصى بدراسة تقدير معالم بواسون في مجالات العلوم الطبية والهندسية والاجتماعية .

المصادر العربية :

- (1) حسن,تارا احمد و رضا,مهدي صابر, "إستخدام الانحدار الحصين لإيجاد أنسب نموذج لتمثل بيانات الأنواء الجوية في مدينة أربيل خلال الفترة (1998-2010)",كلية الادارة والاقتصاد,جامعة صلاح الدين,مجلة الإدارة والاقتصاد , (المجلد-34), (العدد-89), (لسنة 2011) .
- (2) ثامر,مائل كامل, "مقارنة ما بين بعض الطرائق الحصينة وغير الحصينة لتقدير نماذج الانحدار",كلية الإدارة والاقتصاد ,جامعة الانبار,مجلة كلية المعارف الجامعة, (المجلد-30), (العدد-1), (لسنة 2020) .
- (3) ردم,سارة عدنان , "مقارنة بين أنموذج انحدار بواسون وأنموذج ثنائي الحدين السالب مع تطبيق عملي",كلية الإدارة والاقتصاد ,جامعة بغداد,مجلة كلية التراث الجامعة, (المجلد-1), (العدد-37), (لسنة 2023) .

المصادر الاجنبية :

- 4) Wedderburn, R. W.(1974)," Quasi-likelihood functions, generalized linear models,and the Gauss-Newton method", Biometrika,61(3),pp. 439-447 .
- 5) McCullagh, P., and Nelder, J. A.(1989)," Generalized Linear Models" , New York: Chapman & Hall .
- 6) Heyde,C.C.(1997)," Quasi-Likelihood and its Application",Springer-Verlag, New York USA .
- 7) Cantoni,E. and Ronchetti,E.(2001)," Robust Inference for Generalized Linear Models", Journal of the American Statistical Association,96 (455),pp. 1022-1030 .
- 8) Dobson,A. J. (2002)," An introduction to generalized linear models", 2nd edition 2002,Chapman & Hall/CRC, Boca Raton .
- 9) Ver Hoef, J.M. and Boveng, P.L.(2007),"Quasi-Poisson Vs. Negative Binomial Regression: How Should We Model Overdispersed Count Data?", Ecology .
- 10) Winkelmann,R.(2008),"Econometric Analysis of Count Data", 5th Edition, Springer Verlag, Berlin .
- 11) Hosseinian, S.(2009),"Robust Inference for Generalized Linear Models: Binary and Poisson Regression", Doctoral thesis, EPFL Lausanne,Switzerland .
- 12) Hilbe, J. M.(2011),"Negative Binomial Regression", 2nd Edition, Cambridge University Press, Cambridge UK .
- 13) Hosseinian, S. and Morgenthaler, S.(2011)," Weighted Maximum Likelihood Estimates in Poisson Regression". International Conference on Robust Statistics, A Universidade de Valladolid,Italy .
- 14) Lieven Clement,2019-11-08,"Poisson IRWLS", <https://statomics.github.io/SGA2019/assets/poissonIRWLS-implemented.html> .
- 15) Abonazel M. R.(2019),"Advanced statistical techniques using R: outliers and missing data", presented at the 54th Annual Conference on Statistics, Computer Sciences and Operations Research,Cairo University, Egypt .
- 16) Gabriella, A., Abdullah, S. & Soemartojo, S.(2019),"Comparison of poisson and quasi-poisson regression: A simulation study",Proceeding of IEEE International Conference on Software Architecture. pp. 1-8, <https://doi.org/10.29244/icsa.2019.pp1-8> .
- 17) Yuan Zhang's ,(2019),"PubH8401 Linear Models",UNIVERSITY OF MINNESOTA,China,pp. 1-43 .
- 18) Abonazel, M. R. and Saber, O. M.(2020),"A Comparative Study of Robust Estimators for Poisson Regression Model with Outliers", Journal of Statistics Applications & Probability,9(2), pp. 279-286 .

- 19) Abonazel M.R., El-sayed S.M., Saber O.M.(2021),"Performance of robust count regression estimators in the case of overdispersion, zero inflated, and outliers: simulation study and application to German health data". Commun Math Biol Neurosci. 2021:Article ID 55.pp. 1-23, <https://doi.org/10.28919/cmbn/5658> .
- 20) Hartono P. G., Tinungki G. M., Jakaria J., Hartono A. B., Hartono P. G., and Wijaya R. (2021), "Overcoming overdispersion on direct mathematics learning model using the quasi poisson regression",in 1st International Conference on Mathematics and Mathematics Education (ICMMEd 2020) , pp. 442–449.17:29 .
- 21) Sohel Rana, Abu Sayed Md. Al Mamun , Arifur Rahman F. M. , and Hanaa Elgohari, (2023), "Outliers as a Source of Overdispersion in Poisson Regression Modelling: Evidence from Simulation and Real Data",Dept. of Statistics, Univ. of Rajshahi, Bangladesh,International Journal of Statistical Sciences, 23(2),pp. 31-37 .
- 22) Mohammed Sabri Al-zubaidi,Ebtisam k.(2024),"Comparison of Estimation Methods for Zero Truncated Poisson Regression Model",Department of Statistics College of Administration and Economics University of Baghdad,Baghdad,Iraq,Journal of Economics and Administrative Sciences (JEAS), 30(142),pp. 492-508 .

الذكاء الشخصي لدى طلبة قسم الرياضيات وعلاقته بالبخل المعرفي لديهم

أ.د. عبدالواحد حميد الكبيسي/جامعة المعارف/العراق
أ.م.د. فائدة ياسين البديري/ جامعة سامراء/العراق

الذكاء الشخصي لدى طلبة قسم الرياضيات وعلاقته بالبخل المعرفي لديهم

أ.د. عبدالواحد حميد الكبيسي/جامعة المعارف/العراق

أ.م.د. فائدة ياسين البديري/ جامعة سامراء/العراق

ملخص البحث

هدف البحث الكشف عن طبيعة العلاقة بين أحد أهم مفاهيم علم النفس الايجابي المهمة وهو الذكاء الشخصي مع احد المتغيرات شخصية الفرد في الجانب العقلي وهو البخل المعرفي، واتبع المنهج الوصفي التحليلي، واعد مقياس للذكاء الشخصي تكون من 20 فقرة، واختبار للبخل المعرفي وطبق على عينة بلغت (64) طالباً وطالبة من مرحلتين الثاني والرابع من قسم الرياضيات من كلية التربية للعلوم الانسانية، وبينت النتائج: وجود مستوى متوسط من الذكاء الشخصي لدى طلبة قسم الرياضيات. مع وجود مستوى فوق المتوسط من البخل المعرفي. مع وجود فروق في متغير الذكاء الشخصي لصالح الذكور، ووجود فروق في البخل المعرفي لصالح الاناث، وتوجد علاقة ارتباطية عكسية بين الذكاء الشخصي والبخل المعرفي.

Abstract

The aim of the research is to reveal the nature of the relationship between one of the most important concepts of positive psychology, which is personal intelligence, and one of the variables of the individual's personality in the mental aspect, which is cognitive stinginess. The descriptive analytical approach was followed, and a scale of personal intelligence was prepared consisting of 20 paragraphs, and a test of cognitive stinginess was applied to a sample of (64) male and female students from the second and fourth stages of the Department of Mathematics from the College of Education for Human Sciences. The results showed: There is an average level of personal intelligence among students of the Mathematics Department. With an above-average level of cognitive stinginess. With differences in the variable of personal intelligence in favor of males, and differences in cognitive stinginess in favor of females, and there is an inverse correlation between personal intelligence and cognitive stinginess.

المقدمة

تعد مرحلة الدراسة الجامعية من مراحل الحياة أكثرها أهمية، لما لها من أدوار رئيسة في صقل شخصية الطلبة، وتحديد مستقبلهم المهني، بالإضافة إلى تزويدهم كم كبير من المهارات العلمية والعملية والشخصية، وتهدف مرحلة الدراسة الجامعية تنمية مواهب الطلبة وقدراتهم وغرس قيمة الاعتماد على النفس، وبخاصة في مجال تحصيل المعرفة ومجال الحياة العملية، وتزويد المجتمع بالكوادر البشرية في مختلف المرحلة. كذلك تزويدهم بمهارات البحث والاستكشاف والتعلم الذاتي وتنظيم وفهم المشاعر الذاتية (الحسيني & محمد، 2020: 9).

لذا اتجهت الدراسات نحو الاهتمام بقدرات الذكاء اللازمة لفهم وتنظيم المشاعر الذاتية وفهم مشاعر وعواطف الآخرين وهو ما يطلق عليه بالذكاء الشخصي، ويعني معرفة الذات والقدرة على التصرف والوعي بالحالات المزاجية والقدرة على ضبط الذاتي وحب العمل الفردي ويشبه كاردنر الذكاء الشخصي بالتكيف الناجح اتجاه مهام الحياة وتكوين المفاهيم وترتيب المجالات المعرفية (رجب، 2024: 126)

إن الطلبة الذين يتفوقون في الذكاء الشخصي يتمتعون بالقدرة على تشكيل نموذج صادق عن الذات واستخدام هذه القدرة بفاعلية في الحياة بإحساس قوي بالأنس وحبذون العمل منفردين، ولهم إحساسات قوية بقدراتهم الذاتية ومهاراتهم الشخصية، ومن المعلوم أن كل إنسان منا له منسوب معين من الذكاء يقابله منسوب مكمّل من المنطق الفكري، وهذا القدر من الذكاء هو الذي يمنح الإنسان القدرة على التفكير وفهم الأمور الغامضة والمبهمة بالنسبة له. ونتيجة لاختلاف معدلات الذكاء بين الأشخاص، نجد أن بعض الطلبة لا يستطيعون تحديد الحل أو الخيار الأمثل لحل المشكلة أو اتخاذ قرار معين، فنرى البعض يسارع على أي حل يخطر بباليه دون السعي لتحسين قدراتهم العقلية ودراسة الخيارات المطروحة لاتخاذ القرار المناسب والامتنل لحل المشكلة أو تبرير موقف معين (ابوزيد، 2021: 249). فعلى سبيل المثال عندما نوجه بعض الاسئلة للطلبة قسم الرياضيات:

- هناك اشهر في السنة الميلادية فيه (31) يوم وبعض منها فيه (30) يوم اي الاشهر يحتوي على (28) يوم؟ اجاب اغلب الطلبة بانه شهر شباط (فبراير)، في حين كل اشهر السنة تحتوي على 28 يوم.
- لدينا حوض من الزهور يتضاعف يوميا ، اي 2 تصبح 4 في اليوم الثاني و 8 في الثالث وهكذا اكتمل الحوض في اليوم 20 في اي يوم امتلاء نصف الحوض؟ اجاب اغلب الطلبة في اليوم العاشر في حين الجواب الصحيح هو 19.
- نستنتج مما ان اغلب الطلبة يجيب عن سؤال او مشكلة معينة أول إجابة تتبادر إلى اذهانهم وعدم التروي ويقدم الإجابة البديهية ويطلق علماء النفس هذه الظاهرة "البخل المعرفي".
- اي يندفع معظم الناس ونخدع بالأفكار الغير صحيحة، وقد تكون لديهم قدرات ذهنية ومعرفية عالية، لكنهم يتكاسلون عن استخدامها، ولقياس مدى سهولة الانخداع بالأفكار الزائفة، استعانت دراسة (Pennycook, et al.2020:771) باختبار التفكير المعرفي، حيث توصلت ان الافراد الذين حصلوا على درجات متدنية كانوا اضعف قدرة على تمييز الأخبار الزائفة والبخل المعرفي لديهم مرتفع.
- مشكلة البحث
- إن الافراد الذين يتمتعون بدرجة عالية من الذكاء الشخصي يقومون بحل المشكلات المتعلقة بالشخصية بمرونة وتروي وتوافقاتها مع الآخرين بشكل أفضل من غيرهم، وهم قادرون على تحديد أدوارهم الاجتماعية ، وفهم شخصيات الآخرين ويضعون أهدافا منطقية لحياتهم و قدرتهم على حل المشكلات التي تواجههم في مراحل حياتهم و يجعل عملية تطوير نقاط القوة ومعالجة وتقوية نقاط الضعف لدى الفرد وتجعله أكثر مرونة في التفكير، ويعد الذكاء الشخصي الذكاء الرئيسي مقارنة بأنواع الذكاء الاخرى ، وان ارتفاع الذكاء الشخصي لدى الفرد يؤدي الى نجاحه في مجالات حياته من خلال معرفته بمواطن قوته يجعله يختار الحالات التي يقوم بها بشكل جيد او اتخاذ القرار المناسب ، كما ان معرفته بمواطن ضعفه تجنبه الحالات التي يفشل فيها ، وبهذا يعمل على تحسين مواطن ضعفه وتحويلها الى مواطن قوة (Mayer, 2014:49).
- وفق لآراء علماء النفس، فإن العقل البشري قد يميل إلى التفكير وحل المشكلات بطرائق بسيطة والتي لا تتطلب طاقة مرتفعة أو جهداً كبيراً أو وقتاً طويلاً بدلاً من لجوئه إلى طرائق الأكثر تعقيداً وموائمة للظروف، اي أنّ الدماغ يميل إلى البحث عن حلول للمشكلات التي تواجهه بشرط أن تتطلب جهداً عقلياً أقل ، بغض النظر عن مدى حكمته أو ذكائه. حيث ايدت بعض الدراسات المحلية هناك بخل معرفي لدى طلبة الجامعة كدراسة (سليم، 2022)، كذلك اشارة دراسة (قاسم & موسى، 2022) ان طالبات من جامعتين في قسم رياض الأطفال لديهم بخل معرفي مرتفع، وايدت دراسة (علي & حسين، 2022) بان طلبة الجامعة لديهم بخل معرفي مرتفع وان هناك علاقة ارتباطية بين البخل المعرفي والتفكير الحدسي لديهم.
- وبهذا تصاغ مشكلة البحث بالسؤال الاتي : هل توجد علاقة بين الذكاء الشخصي والبخل المعرفي لدى طلبة قسم الرياضيات ؟.

اهمية البحث

ويمكن تلخيص اهمية الدراسة بجانبين:

(أ): الأهمية النظرية:

- يتناول هذا البحث شريحة مهمة من المجتمع تتمثل في طلبة الجامعة الذين يعدون ركيزة المجتمع وأساس تقدمه وازدهاره، ويعد مساهمة متواضعة في إثراء المكتبة التربوية والنفسية بما يخدم الباحثين في هذا المجال في المستقبل.
- زيادة وعي الطلبة الجامعيين بمدى اهمية الذكاء الشخصي والبخل المعرفي في حياتهم.

(ب) ثانياً: الأهمية التطبيقية:

يوفر البحث الحالي أداتين لقياس (الذكاء الشخصي، والبخل المعرفي) لدى طلبة الجامعة يمكن الاستفادة منه في البحوث العلمية مستقبلاً.

اهداف البحث:

يهدف البحث الحالي الى التعرف على:

1. مستوى الذكاء الشخصي لدى طلبة قسم الرياضيات.
2. مستوى البخل المعرفي لدى طلبة قسم الرياضيات.
3. العلاقة بين الذكاء الشخصي والبخل المعرفي
4. الفروق ذات الدلالة الإحصائية في الذكاء الشخصي لدى طلبة قسم الرياضيات تبعاً لمتغيري الجنس والمرحلة .
5. الفروق ذات الدلالة الإحصائية في البخل المعرفي لدى طلبة الجامعة تبعاً لمتغيري الجنس والمرحلة .

حدود البحث:

يتحدد البحث الحالي بطلبة قسم الرياضيات في كلية التربية للعلوم الصرفة في جامعة الانبار للدراسة الصباحية المرحلة الثانية والرابعة للعام الدراسي (2023 – 2024).
تحديد المصطلحات اجرائياً:

- الذكاء الشخصي: قدرة الطالب الجامعي على ادراك نقاط القوة والضعف لديه وامكانية تعلمه بالقدر وبالأسلوب المناسب، ويقاس بالمقياس المعد في البحث.
- البخل المعرفي: اعتماد الطالب الجامعي على الحلول العقلية السريعة عند التعامل مع المواقف او الاسئلة التي تحتاج تحليلات مجهزة ويقاس بالاختبار المعد في البحث.

الفصل الثاني: اطار نظري ودراسات سابقة

المحور الاول: الذكاء الشخصي

الذكاء الشخصي: هو قدرة الفرد على معرفة عالمه (الداخلي، والخارجي) ومشاعر حياته ومدى انفعالاته والتصرف بصورة تكيفية على أساس تلك المعرفة داخلياً (ذاتياً) يشمل معرفة وفهم الذات، الاستقلال في التفكير، والعمل، تقدير الذات، وتكوين نموذج عنها، اما خارجياً (اجتماعياً) فيشمل الإحساس والتعاطف مع الآخرين، التفاعل والتعاون مع الآخرين، التأثير في الآخرين (عبدالرؤف & عيسى، 2018: 154).

الذكاء الشخصي أحد انواع الذكاءات المتعددة لنظرية العالم كاردنر (1983)، الذكاء الشخصي - المصطلح مواز للذكاء العاطفي والاجتماعي، هو قدرة عقلية تنطوي على التفكير الدقيق حول الشخصية في الذات وفي الآخرين. ويشمل ذلك القدرة على مراقبة المعلومات ذات الصلة بالشخصية، اما مجالات الذكاء الشخصي (الداخلي (الذاتي) وهي:-

الأول: معرفة الذات وتعني القدرة على التعرف بصورة تكيفية لذاته وعلى هذا الاساس وتلك المعرفة تتبلور نظرة الامتلاك صورة دقيقة عن الذات لتشمل نقاط القوة ونقاط الضعف فيه لذاته..

الثاني: الاستقلالية في التفكير والعمل وهي القدرة التي يعتمد الفرد فيها على نفسه فيها ويتحمل مسؤولية أعماله ومواجهة الضغوط الحياتية والاجتماعية.

الثالث: تقدير الذات وتكوين نموذج عن الذات ذو نسق معرفي ووجداني يكونه الفرد من خلال مجموعة من المعتقدات والاتجاهات والخصائص العقلية والحسية التي يستدعيها الفرد عندما يواجه العالم المحيط به ويستطيع تكوين نظرة إيجابية عن نفسه واحساسه بكفاءته الذاتية

(Allen, & Mayer, 2022: 1)

ب- أما فيما يخص الذكاء الشخصي الخارجي (الاجتماعي) فيذكر كاردنر أنه يتجه إلى الخارج إلى الأفراد الآخرين، والفكرة الاساسية هنا هي القدرة على ملاحظة التمييزات بين الأشخاص والقيام بها خصوصاً أمزجتهم وطبائعهم ودوافعهم ومقاصدهم، وأكثر صورة الذكاء بدائية هي قدرة الطفل الصغير على التمييز بين الأفراد من حوله وتحديد أمزجتهم المتنوعة، أما في صورته المتقدمة فإن يعني أن القدرة على المعرفة للأشخاص والتي تتيح للراشد الحاذق قراءة مقاصد الناس ورغباتهم حتى لو كانت خافية عن أفراد آخرين كثيرين ثم امتلاك القدرة على الإحساس والتصرف على وفق هذه المعرفة. (Teodorescu, 2021: 117).

لذا نرى اهمية دراسة الذكاء الشخصي لدى طلبة الجامعة كون الحياة الجامعية بأنها فترة حساسة وحاسمة في حياتهم تتسم ببناء القدرات العقلية و تطوير المهارات الأدائية المستقبلية اللازمة، أي يعد الذكاء الشخصي من العوامل الحاسمة لنجاحهم، إذ إن الطلبة من ذوي الذكاء المرتفع يتسمون بالتنوع في ميولهم المهنية والعلمية بدرجة أكبر من الطلبة ذوي الذكاء المنخفض، ويأثر الذكاء الشخصي أيضاً على إدراك الطلبة لنقاط القوة والضعف لديهم، وإمكانية التعلم بأساليب مناسبة (Gardner, 2013: 44)، كما وينعكس المستوى المتدني للذكاء الشخصي لدى طلبة الجامعة بعدم وجود الانجاز الشخصي، مما قد يؤثر سلباً على تحقيق الأهداف الأكاديمية، ويصعب عليهم التطوير والتقدم العلمي والمهني والنفسية، إذ توصلت دراسة (الحيار، 2020) بوجود علاقة ارتباطية موجبة دالة إحصائياً بين درجات طلبة الجامعة بين الذكاء الشخصي و دافعية الإنجاز الأكاديمي. وبينت دراسة (جعفر، 2020): وجود علاقة بين الذكاء الشخصي والتعاطف الاجتماعي لدى طلبة الجامعة.

واظهرت دراسة (Ibrahim & Abu Hammad, 2021) العلاقة الارتباطية بين الذكاء الشخصي الاجتماعي والمسؤولية الاجتماعية والتوافق كما يراها معلوم المرحلة الثانوية. كذلك بينت دراسة (عمران 2023) بوجود علاقة ارتباطية ايجابية بين الذكاء الشخصي والصمود

النفسي.

وبينت دراسة (Bryan, 2023): دور الذكاء الشخصي القائم على القدرة على التواصل في الصراعات الاجتماعية المحور الثاني: البخل المعرفي

مصطلح البخل المعرفي ظهر من ضمن مفاهيم علم النفس المعرفي لأول مرة عام 1984 م من قبل عالمي النفس الأمريكيين (Suzan Fiske) ، (Shelly Taylor) ، ووصفا البخل المعرفي بأنه الطريقة التي يستجيب فيها الطلبة لحل مشكلة ما أو استجابة للمواقف التي تواجههم، وتكون تلك الاستجابة من نوع نمط التفكير البسيط الذي يتصف بالسرعة ، فالبخل معرفياً يتجنب التفكير بالأساليب التي تحتاج إلى متابعة مستمرة و طاقة عقلية كبيرة ووقت وجهد وتفكير وتحليل عميق (Svedholm & Kiikeri, 2022: 333).

على سبيل المثال؛ طرحنا السؤال الاتي سيكون عُمر حسن في غضون ثلاثة اعوام ضعف عمره منذ سبع سنوات، فكم عمر حسن الآن، كانت اجابات اغلب الطلبة 14 سنة وقسم اجاب 10 سنوات و اجاب طالب واحد بقوله اعمل معادله ونفرض عمر حسن=س، فتكون المعادلة:

$$س + 3 = 2(س - 7) \text{ ومنها } س + 3 = 2س - 14 \text{ وبحل المعادلة ينتج } س = 17 \text{ سنة}$$

أي الطلبة البخلاء معرفياً يقدمون استجابة سريعة لأول فكرة تخطر فيا ذهنهم، وتكون اجاباتهم خاطئة. بعض الطلبة يعتمدون على استراتيجية الاستدلال العقلي السريع ، ويضنون ان الاجابة السريعة تكون أكثر وظيفية في مثل بعض المواقف، الا أنها تدل على ارتفاع مستوى البخل المعرفي لديهم ، وعدم الاهتمام بانفعالاتهم ومراقبة ما وراءها أثناء التعامل مع بعض المواقف ، وبهذا النمط تكون المعالجة بخيلة لمعالجة المعلومات و يؤدي إلى استجابات دون المستوى المطلوب ، من خلال فشل الطالب في السيطرة على انفعالاته، مما يؤدي إلى فشله في منع ميله الطبيعي للبخل المعرفي ، و يستطيع السيطرة على العمليات الحسية ، ولا يتمكن من منعها في التعامل مع المهمة التي تواجهه، البخل المعرفي يدل على أن لدى الطالب معلومات ما يميل إلى البخل حفاضاً من النفاذ ، ألا وهو التنظيم الذاتي ، وهو أمراً حيوياً يحتاجه الفرد للغاية لإكمال الكثير من المهمات خاصة تلك التي تكون أكثر اثارة من الانفعالات ، و يصبح الطالب أكثر بخلأ معرفياً نفدت موارده في التنظيم الذاتي لديه، فيندفع أكثر نحو استخدام استراتيجيات البخل المعرفي (Stanovich, 2018: 423).

وهذا ما ايدته دراسة (Vonasch, 2016) التي كشفت عن وجود علاقة بين نضوب التنظيم الذاتي و البخل المعرفي. و دراسة (العظامات، حسابان، 2021) التي بينت للتنظيم الذاتي قدرة في تنبؤ بالبخل المعرفي لدى طلبة الجامعة. اما دراسة (حمودة، 2021) بينت علاقة ارتباطية عالية بين التضليل المعرفي والبخل المعرفي.

الفصل الثالث

اجراءات البحث

منهج البحث: اتبع منهج الوصفي التحليلي الارتباطي

مجتمع البحث: تكون المجتمع من طلبة قسم الرياضيات في كلية التربية للعلوم الصرفة من جامعة الانبار المرحلة الثانية بواقع (71) طالب وطالبة للدراسة الصباحية والرابعة (141) طالب وطالبة عينة البحث: تم اخذ نسبة 30% من مجتمع البحث وسرت هذه النسبة على مكونات مجتمع البحث وبلغت 64 طالب وطالبة منهم 21 طالب وطالبة من الصف الثاني و (43) طالب وطالبة من الصف الرابع. ادوات البحث: تمثلت ادوات البحث:

اولاً: مقياس الذكاء الشخصي: بعد مراجعة صياغة الفقرات من الأدبيات والدراسات السابقة والمقاييس ذات العلاقة بموضوع الذكاء الشخصي وقد بلغ عدد الفقرات بصيغتها الأولية (29) فقرة ، مقابل مقياس ثلاثي لدرجة الانطباق (عالية، متوسطة، منخفضة)، موزعة على اربع مجالات (معرفة وفهم الذات، الاستقلال والتفكير في العمل، تقدير الذات وتوجيهها، تكوين انموذج عن الذات)، وقد روعي في صياغتها أن تكون قصيرة وذات لغة مفهومة لعينة البحث . ولتقرير مدى صلاحية الفقرات في مقياس الذكاء الشخصي التي أعدت لقياسه ، عرضت الفقرات على (12) خبيراً من المتخصصين في العلوم التربوية والنفسية ، وطلب منهم تقدير مدى قدرة كل فقرة كما تبدو ظاهرياً في قياس المجال الفرعي التي أعدت لقياسه وبناءً على آرائهم وملاحظاتهم عدلت بعض الفقرات وحذفت (2) فقرة لأنها لم تحصل على موافقة (80%) من الخبراء وبقيت 27 فقرة. و للتأكد من وضوح فقرات المقياس وتعليماته طبق على عينة استطلاعية ممثلة لمجتمع البحث عددهم (18) طالباً وطالبة من غير عينة البحث الاصلية، لبيان فيما اذا كانت الفقرات واضحة ، وتبين ان الفقرات وتعليماتها واضحة وتم حساب الوقت الازم للمقياس بحساب مجموع زمن انتهاء كل طالب وطالبة مقسوماً على العدد $\frac{217}{18} \approx 12$ دقيقة.

وتم تطبيق المقياس مرة أخرى على عينة (60) طالباً وطالبة من غير عينة البحث لتحليل الاحصائي لكون مجتمع البحث قليل وذلك بقسمتهم الى مجموعتين لتمثلان المجموعة العليا والدنيا، وتم حساب إيجاد القوة التمييزية لفقرات المقياس بتطبيق (T-Test) لعينتين مستقلتين لاختبار دلالة الفروق بين المجموعتين العليا والدنيا لكل فقرة من فقرات المقياس. وقد عدت القيمة الثانية مؤشراً على تمييز كل فقرة من خلال مقارنتها بالقيمة الثانية الجدولية. وتبين أن فقرات المقياس جميعها مميزة عند مستوى دلالة (0.05)، ودرجة حرية (59)، باستثناء فقرة واحدة تم استبعادها لتبقى 26 فقرة.

وتم للمقياس صدق الاتساق الداخلي من خلال ارتباط درجة كل فقرة بالدرجة الكلية للمقياس: وظهر أن جميع معاملات الارتباط المحسوبة دالة إحصائياً عند مقارنتها بالقيمة الجدولية ما عدا فقرة واحدة وتم استبعادها أيضاً من المقياس لتبقى 25 فقرة للمقياس بصورة نهائية.

وتم تطبيق الثبات من خلال تطبيق المقياس على المجموعة الاستطلاعية (18) طالب وطالبة ثم تصحيحه وتدوين نتائجه، وبعد مدة زمنية (أسبوعين) يعاد تطبيق المقياس على نفس العينة وضمن ظروف مشابهة وتصحيحه وتدوين نتائجه مرة ثانية، ثم حساب معامل الارتباط بين درجات التطبيقين، ويسمى الارتباط المستخرج بهذه الطريقة (معامل الاستقرار) والذي يبلغ (0.89). وبذلك يكون المقياس جاهزاً للتطبيق مكون من 25 فقرة وأعلى درجة (75) وأقل درجة (25) والوسط الفرضي بلغ (50) درجة.

ثانياً: اختبار البخل المعرفي

تم اعتماد اختبار (Toplak, et al. 2014) والمكون من 12 فقرة تم ترجمتها وتكييفها بما يتناسب والبيئة العراقية وتم عرضه على مجموعة الخبراء وحصلت الموافقة على جميع الفقرات بنسبة 100%.

للتأكد من وضوح فقرات الاختبار وتعليماته طبق على عينة استطلاعية السابق ذكرهم وعددهم (18) طالباً وطالبة، لبيان فيما إذا كانت الفقرات واضحة، وتبين أن الفقرات وتعليماتها واضحة وتم حساب الوقت للإجابة العدد $\frac{723}{18} \approx 40$ دقيقة.

وتم تطبيق المقياس مرة أخرى على نفس العينة (60) طالباً وطالبة وبتطبيق (T-Test) لعينتين مستقلتين لاختبار دلالة الفروق بين المجموعتين العليا والدنيا لكل فقرة. وتبين أن فقرات الاختبار جميعها مميزة.

وتم للمقياس صدق الاتساق الداخلي من خلال ارتباط درجة كل فقرة بالدرجة الكلية للاختبار: وظهر أن جميع معاملات الارتباط المحسوبة دالة إحصائياً عند مقارنتها بالقيمة الجدولية ما عدا لتبقى 9 فقرات للاختبار بصورة نهائية.

وتم تطبيق الثبات من خلال تطبيق المقياس على المجموعة الاستطلاعية (18) طالب وطالبة ثم تصحيحه وتدوين نتائجه، وبعد مدة زمنية (أسبوعين) يعاد تطبيق الاختبار وبلغ (0.85). وبذلك يكون الاختبار جاهزاً للتطبيق موزعة على أربعة ابعاد (استبدال العزو- الاختصارات العقلية- الاستدلال العقلي السريع- الاستدلال الانفعالي)، أعلى درجة 27 وأقل درجة 9 والمتوسط الفرضي 18 بحيث الإجابة (الصحية) (1)، حدية إجابة خاطئة (3)، أخرى غير حدية خاطئة (2).

نتائج البحث

أولاً نتائج الذكاء الشخصي

بعد تصحيح إجابات العينة على مقياس الذكاء الشخصي يبين الجدول (1) نتائج ذلك:-

المقياس	العدد	الوسط الحسابي	الانحراف المعياري	الوسط الفرضي	القيمة الثانية		درجة الحرية	الدلالة الاحصائية
					المحسوبة	الجدولية		
الذكاء الشخصي	64	57.34	6.18	50	9.501618	2.00	63	دال عند 0.05

من خلال الجدول يتبين أن الفروق داله احصائيا وان طلبة الكلية يتمتعون بذكاء شخصي بدرجة متوسطة، وهذا يدل ان فترة دراستهم في الجامعة تجعلهم مدركون لعالمهم الداخلي وحدود قدراتهم وقادريين على التكيف مع البيئة الجامعية، من خلال معرفة حالاتهم النفسية الداخلية، والدوافع والنوايا والأمزجة والمشاعر والرغبات، وفهم الذات وتقديرها، وتفسر هذه النتيجة إلى صفات شخصية طلبة الجامعة من خلال مواجهتهم للتحديات العديدة الأكاديمية والاجتماعية وحتى المادية خلال فترة الدراسة، مما طور من ذكائهم الشخصي، وقدرتهم على التحليل العقلي، واستخدام المهارات المعرفية

وللتعرف فيما إذا كانت الفروق في متوسطات الذكاء الشخصي على وفق متغير الجنس ذي دلالة احصائية استخدم الاختبار التائي لمجموعتين مستقلتين غير متساويتين بالعدد يبين الجدول (2) ذلك

المجموعة	العدد	المتوسط الحسابي	التباين	درجة الحرية	t-test		الدلالة الإحصائية عند (0.05)
					المحسوبة	الجدولية	
الطالبات	21	59.11	21.17	62	3.812942	2.00	دال احصائياً
الطلاب	43	55.57	31.52				

يتبين من الجدول (2) وجود فروق بين إجابات عينة البحث تبعاً لمتغير الجنس ولصاح الذكور. وقد تبدو النتيجة منطقية على وفق المشاهدات الحياتية التي نلاحظها يومياً، حيث نجد أن الذكور أعلى انفتاحاً على البيئة المحيطة بهم مقارنة بالإناث؛ نظراً لطبيعة التنشئة الاجتماعية التي يحدد الانفتاح على البيئة المحيطة بهم ويكتسبوا علاقات مجتمعية متنوعة. وقد يكون لدى الذكور خبرات يومية أكثر من الإناث، وأنهم قد يمرون بتجارب مستمرة التي تفرضها عليهم طبيعتهم، والظروف التي يمرون بها قد تفرض عليهم خوض التجارب، مما يكسبهم مرونة أكثر مقارنة بالإناث في تقبل التغيرات المحيطة بهم. وللتعرف فيما إذا كانت الفروق في متوسطات الذكاء الشخصي على وفق متغير المرحلة (الثاني- الرابع) ذي دلالة احصائية استخدم الاختبار التائي لمجموعتين مستقلتين غير متساويتين بالعدد يبين الجدول (3) ذلك

المجموعة	العدد	المتوسط الحسابي	التباين	درجة الحرية	t-test		الدلالة الإحصائية عند (0.05)
					المحسوبة	الجدولية	
الرابع	43	58.51	30.26	62	2.481916	2.00	دال احصائياً
الثاني	21	56.17	26.43				

يتبين من الجدول وجود فروق بين إجابات عينة البحث تبعاً لمتغير المرحلة حول الذكاء الشخصي ولصالح المرحلة الرابعة، وقد يعزى سبب التفوق كون المرحلة الرابعة تم اكتسابهم معلومات علمية وتربوية خلال دراستهم الجامعية، و يكون لديهم قدر عالٍ جداً من مهارات الانتباه المركز بحكم تخصصهم التي تتطلب منهم ضرورة التركيز على فكرة معينة حتى انتهاء المهمة الدراسية في موضوع معين أو المواقف التي تتطلب حلاً، ونجدهم أكثر انفتاحاً وأكثر بحثاً عن الجديد بحكم التطور السريع في مجال تخصصهم، لذا فإنهم يكونون أكثر مرونة، وقدرة على إجراء مناورات مختلفة في أثناء التعامل مع مهام دراستهم ويمتلكوا القدرة السريعة على تغيير تكتيك الحل للمواضيع والمسائل العلمية، والعمل في المهمة كلما لاحظوا أي تغير في معطياتها، وهذا من شأنه قد يؤدي إلى إحداث نوع من التغير في المهارات والقدرات الشخصية والاجتماعية التي تعد من أساسيات للذكاء الشخصي مما يكسبهم درجة أعلى في الذكاء الشخصي مقارنة بالمرحلة الثانية التي لم تكتمل تخصصهم العلمي والتربوي.

النتائج الخاصة بالبخل المعرفي

بعد تصحيح إجابات العينة على اختبار البخل المعرفي يبين الجدول (4) نتائج ذلك:-

المقياس	العدد	المتوسط الحسابي	الانحراف المعياري	الوسط الفرضي	القيمة التائية		درجة الحرية	الدلالة الاحصائية
					المحسوبة	الجدولية		
الذكاء الشخصي	64	21.33	4.148	18	6.422372	2.00	63	دال عند 0.05

وللحكم على مستوى البخل المعرفي اعتمد الجدول (5) الاتي وفق تعليمات الاختبار

مستوى البخل	الدرجة	
	إلى	من
منخفض	13	9
اقل من وسط	16	14
متوسط	19	17
اعلى من متوسط	22	20
مرتفع	27	23

من خلال الجدول (4) يتبين ان الفروق داله احصائيا وان طلبة الكلية لديهم بخل معرفي بدرجة ووفق الجدول (5) لدى الطلبة بخل معرفي بمستوى اعلى من متوسط ويمكن تفسير تلك النتيجة ، أن أغلبية الطلبة يميلون إلى الاستجابات الحدية القائمة على الاستدلالات العقلية السريعة ، وعلى الاستدلالات الانفعالية ، التي تشعرهم بسهولة الموقف رغم ما يبدو عليه من صعوبة ، كنوع من التشجيع لأنفسهم ؛ ليخفف عليهم تبعات الشعور باحتمال الفشل ، وكذلك يميلون إلى الاختصارات العقلية ، رغم كونهم تخصص رياضيات.

وللتعرف فيما إذا كانت الفروق في متوسطات الذكاء الشخصي على وفق متغير الجنس ذي دلالة احصائية استخدم الاختبار التائي لمجموعتين مستقلتين غير متساويتين بالعدد يبين الجدول (6) ذلك

المجموعة	العدد	المتوسط الحسابي	التباين	درجة الحرية	t-test		الدلالة الإحصائية عند (0.05)
					المحسوبة	الجدولية	
الطالبات	21	23.32	12.25	63	5.499695	2.00	دال احصائياً
الطلاب	43	19.34	19.45				

يتبين من الجدول (6) ان الفروق في البخل المعرفي وفق متغير الجنس دالة احصائياً ولصالح الطالبات، ويمكن ان تفسر ذلك ان الطلاب يكون لديهم ميل أكبر من الإناث إلى اطفاء الاستجابات الحدية ، وبذل الجهد العقلي بصورة اكبر في المهمات التي يتعرضون لها، ليظهروا أمام الآخرين بانهم أكثر صلابة و قدرة على تحمل المسؤولية ، ومقاومة تبعية انفعالاتهم ، قد يؤدي إلى انخفاض مستوى البخل، اما الطالبات فقد يميلن اعلى الاعتماد على العواطف و المشاعر والانفعالات نحو ما يعرض عليهم من مهمات ، فيتجهن إلى إصدار استجابات انفعالية و استجابات استدلالية سريعة ، حتى يتجنبن بذل مزيد من الجهود العقلية حتى لا يبدين على نحو صارم جاف لا يرغب فيه ، ظنا منهن أن الآخرين لا يفضلون ذلك ويميلون إلى العاطفية التي جبلوا عليها وفق طبيعة خالق البشرية ، وقد يؤدي هذا إلى ارتفاع مستوى البخل المعرفي لديهن.

وللتعرف فيما إذا كانت الفروق في متوسطات البخل المعرفي على وفق متغير المرحلة (الثاني- الرابع) ذي دلالة احصائية استخدم الاختبار التائي لمجموعتين مستقلتين غير متساويتين بالعدد يبين الجدول (7) ذلك

المجموعة	العدد	المتوسط الحسابي	التباين	درجة الحرية	t-test		الدلالة الإحصائية عند (0.05)
					المحسوبة	الجدولية	
الرابع	43	19.79	9.61	62	5.02769	2.00	دال احصائياً
الثاني	21	22.87	17.64				

يتبين من الجدول (7) وجود فروق في مستوى البخل المعرفي بين مرحلتَي الثاني والرابع ولصالح المرحلة الثانية، ونعزوا سبب ذلك ان طالب مرحلة الرابعة قد تم مدة اربع سنوات قد دربتهم على ضرورة إعمال العقل وبذل الجهد العقلي المناسب للتفكير حسب كل مشكلة او موقف ، وعلى ضرورة الاعتماد على الأدلة الفعلية الواقعية وليس على ما يشعر به الطالب من انفعالات تحمها مثيراتها وفق ما توجي به كل مهمة ، لذا فان مستوى البخل المعرفي قد تحسن مع انتقالهم من مرحلة دراسية إلى المرحلة التي تليها ، كما ويبدو منطقياً أنه كلما تقدم الطالب في العمر كلما أصبح أكثر عقلانية وأكثر ميلاً إلى تجنب الانفعالات عند اصدار الحكم على الأمور المحيطة به ، وأكثر ميلاً إلى عدم الانسياق وراء الانفعالات والمشاعر المصاحبة للمواقف حتى لا تعوق تفكيره وتؤثر عليه ، وقد يؤدي هذا إلى انخفاض مستوى البخل المعرفي لديه ، وخاصة المرحلة الرابعة الهامة في حياته التي تتحدد فيها ملامح شخصيته عامة ويأهل للعمل الميداني في التدريس.

ولإيجاد العلاقة بين الذكاء الشخصي والبخل المعرفي طبق قانون معامل ارتباط بيرسون بين درجات كلاهما فوجد انهو يساوي(-0.65) وهو معامل ارتباط عكسي متوسط ويشير إلى أنه كلما ارتفع مستوى الذكاء الشخصي لدى الطالب كلما انخفض مستواه إجمالاً في البخل المعرفي .

الاستنتاجات

من خلال نتائج التي تم التوصل إليها نستنتج الآتي:

- (1) يتمتع طلبة قسم الرياضيات بمستوى متوسط من الذكاء الشخصي.
- (2) يوجد لدى طلبة قسم الرياضيات بخل معرفي بمستوى فوق المتوسط .
- (3) توجد علاقة ارتباطية عكسية متوسطة بين الذكاء الشخصي والبخل المعرفي.
- (4) لمتغير الجنس تأثير في امتلاك الذكاء الشخصي ولصالح الذكور.
- (5) لمتغير المرحلة تأثير في امتلاك الذكاء الشخصي ولصالح المرحلة الأعلى.
- (6) لمتغير الجنس تأثير في البخل المعرفي ولصالح الإناث.
- (7) لمتغير المرحلة تأثير في البخل المعرفي ولصالح المرحلة الأدنى.

التوصيات

من خلال النتائج التي توصل إليها البحث نوصي الآتي:

- 1- على المؤسسات التربوية الجامعية توظيف المناهج الدراسية ببعض المواقف التي تتضمن استثارة التفكير ؛ لتعويدهم وتدريبهم على التعامل مع المشكلات في الحياة العملية .
- 2- عقد دورات تدريبية من قبل الكلية والأقسام ، وبرامج تثقيفية لتنمية الذكاء الشخصي لدى طلابها.
- 3- عقد اقسام الرياضيات نشاطات ومسابقات تتضمن فقرات من اختبار البخل المعرفي بهدف خفضه.
- 4- اقامة ورش لتدريب التفكير ومبادئ إعمال العقل ، والبعد عن الارتجالية والعشوائية عند حل المشكلات واتخاذ القرارات المناسبة في المواقف الحياتية .
- 5- على مدرسي الرياضيات تعويد الطلبة على أن كل نتيجة يصلون إليها ترتبط بسبب منطقي ، وأنه من الضروري بذل الجهود العقلية بشكل يناسب طبيعة مسألة رياضية او موقف معين.

المصادر

- 1- ابو زيد، محمد مبروك.(2021). الفيزياء البشرية: عالمٌ كبير.. تُحرّكه تفاصيلٌ دقيقة وقوانين مفقودة، القاهرة، مصر.
- 2- الجبار، دينا جلال.(2020). الذكاء الشخصي وعلاقته بالدافعية للإنجاز الأكاديمي لدى الطلبة المعلمين بكلية التربية جامعة حلوان، دراسات تربوية واجتماعية - كلية التربية - جامعة حلوان، المجلد(26)، العدد يناير الجزء (4)، ص118- 154.
- 3- الحسيني، فايزة أحمد & محمد، عبدالناصر شريف.(2020). مهارات الطالب الجامعي المتفوق دار التعليم الجامعي، القاهرة، مصر.
- 4- حمودة، عبدالواحد.(2021). الدور الوسيط للتحيزات المعرفية في العلاقة بين التضليل المعرفي حول كورونا والبخل المعرفي والتفكير المنفتح النشط لدى عينة من طلاب الجامعة، المجلة المصرية للدراسات النفسية، المجلد(31)، العدد(110)، ص131- 210.
- 5- رجب، أحمد.(2024). تشكيل الأنا.. أسس بناء الشخصية وتطوراتها، كالة الصحافة العربية، القاهرة، مصر.
- 6- سليم، مظهر عبدالكريم.(2022). البخل المعرفي عند طلبة الجامعة، مجلة ديالى للبحوث الانسانية، العدد(91)، ص43- 68.
- 7- عبدالرؤف، طارق & عيسى، ايهاب،.(2018). الذكاء العاطفي والذكاء الاجتماعي، المجموعة العربية للتدريب والنشر، القاهرة، مصر.
- 8- العظامات، عمر عطالله & حسيان، تمار قاسم القدرة التنبؤية للتنظيم الذاتي بالبخل المعرفي لدى طلبة الجامعة، مجلة جامعة الحسين بن طلال للبحوث، المجلد(7)، العدد(2)، ص391- 417.
- 9- عمران، رنا حسين.(2023). الذكاء الشخصي وعلاقته بالصمود النفسي لدى طلبة الجامعة، مجلة آداب المستنصرية، المجلد(47)، العدد(103)، ص393- 420.
- 10- قاسم، زهراء & موسى، ميادة.(2022). قياس البخل المعرف لدى طالبات قسم رياض الاطفال، مجلة نسق، مجلد(35)، العدد(10)، ص591- 610.

المصادر الاجنبية

- 1- Allen, Jayne & Mayer, John D. (2022). Personal Intelligence Is Evident in the Sophistication of People's Narratives about Personality, Journal of Intelligence 10, 56,p.1-20.
- 2- Bryan, Victoria. (2023). "Knowing Me, Knowing You: The Role of Personal Intelligence in Romantic Relationship Conflict Communication" Doctoral Dissertations.
- 3- Ibrahim و Naser & Abu Hammad, Ahmad.(2021). The Personal-Social Intelligence And Its Relationship To Social Responsibility And The Tendency To Keep Up With From The Viewpoint Of Secondary School Teachers, Multicultural Education Volume 7, Issue 3,p.63- 74.
- 4- Mayer, J.D. (2014): personal intelligence the power of personality and how it shapes our lives, New York.
- 5- Pennycook, Gordon,et al.(2020). Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention, Psychological Science, Vol. 31(7) 770–780.
- 6- Stanovich, K.E. (2018). Miserliness in human cognition: the interaction of detection, override and mind ware, J. of Thinking & Reasoning, (24) 4: 423 – 444.
- 7- Svedholm ,Annika & Kiikeri ,Mika .(2022). Cognitive miserliness in argument literacy? Effects of intuitive and analytic thinking on recognizing Fallacies, Judgment and Decision Making, Vol. 17, No. 2, , pp. 331–361
- 8- Teodorescu, Ioana Bucur.(2021). The Impact of the Digital Consumer's Emotional Intelligence in Relation to the Moral Values Promoted in E-Busines ,Springer Nature.
- 9- Toplak, M.E., West, R.F. & Stanovich, K.E. (2014). Assessing miserly information processing: an expansion of the cognitive reflection test, J. of Thinking & Reasoning, (20) 2: 147 – 168
- 10- Vonasch, A.(2016). *Cognitive miserliness preserves the self-regulatory resource*, Unpublished PhD Thesis, Florida: Florida State University.

Estimation of the multiple linear regression model for the production quantity of Al-Zahi in Al-Mamoun Factory using the (r-(k-d)) method in the presence of the problems of autocorrelation and multicollinearity together

**Saja Mohammad Hussein
Zainab abd alsatar**

Baghdad university , college of fine arts, financial Dept.

Baghdad university , college of administration and economic, Dept. of Statistics

zainab.a@cofarts.uobaghdad.edu.iq

Estimation of the multiple linear regression model for the production quantity of Al-Zahi in Al-Mamoun Factory using the (r-(k-d)) method in the presence of the problems of autocorrelation and multicollinearity together

Saja Mohammad Hussein

Zainab abd alsatar

Baghdad university , college of
administration and economic,
Dept. of Statistics
saja@coadec.uobaghdad.edu.iq

Baghdad university , college of fine arts, financial Dept.
zainab.a@cofarts.uobaghdad.edu.iq

Abstract:

The controlling production is one of the important goals that economic institutions strive to achieve. The quantities of production are essential for industrial establishments, and one of the most important statistical models used is the linear regression model. Ordinary Least Squares (OLS) is commonly used to estimate the linear regression model. However, this method requires a set of assumptions. If one or more of these assumptions are violated, it may lead to inaccurate and undesired estimates.

These problems often manifest in the multiple linear regression model as problem of multicollinearity and autocorrelation together. In this study, **r-(k-d)** method was used to estimate the model for the production quantity of cleaning materials in AL-Maamon industry in the General Company for vegetable oil industry. This approach addressed the issues of multicollinearity and autocorrelation simultaneously.

Keywords: Multicollinearity, Autocorrelation, Multiple Linear Regression Model, Generalized Least Squares (GLS).

1-المقدمة وهدف البحث

ان تحليل الانحدار هو احد الاساليب الاحصائية واسعة الاستعمال في تحليل البيانات وذلك لسهولة استعمالها في معرفة وتفسير العلاقات المختلفة بين الظواهر عن طريق تحديد العلاقات بين المتغيرات و يعد تحليل الانحدار الخطي المتعدد احد اكثر الاساليب استعمالا في كثير من المجالات من اجل الوصول الى نتائج افضل للعلاقة المدروسة بواسطة الاستعمال الامثل للبيانات . ويفترض فيه وجود علاقة خطية بين المتغير المعتمد Y (dependent variable) وعدة متغيرات توضيحية X (Explanatory variable) والهدف من تحليل الانحدار الخطي المتعدد هو التنبؤ بقيم مستقبلية وكذلك يستعمل لتقدير قيم سابقة ويكون ذلك بالاعتماد على دراسة حالات اخرى. ولانموذج الانحدار المتعدد الاتي:

$$\underline{Y} = X \underline{\beta} + \underline{U} \quad \dots\dots\dots(1)$$

اذ ان : \underline{Y} : تمثل متجه من الدرجة $n \times 1$ لملاحظات المتغير المعتمد

X : تمثل مصفوفة المتغيرات التوضيحية من الدرجة (p+1) × n

β: تمثل متجه من الدرجة (p+1) × 1

n : تمثل عدد المشاهدات

p : هي عدد المتغيرات التوضيحية

U : حد الخطأ من الدرجة 1 × n

ولكي نتمكن من تقدير معاملات نموذج الانحدار المتعدد (General Linear Model) باستعمال المربعات الصغرى الاعتيادية فينبغي توفر الفروض [2] وعند عدم تحقق احد الفروض في نموذج الانحدار فسوف تظهر مشاكل عديدة فعندما لا تتحقق الفروض الاتية:

ان تكون الاخطاء العشوائية غير مرتبطة خطيا مع بقية الاخطاء العشوائية الاخرى اي ان قيم الخطأ العشوائي في اي فترة زمنية لا تعتمد على قيمته في فترة زمنية اخرى بحيث ان

$$\text{Var-cov} (U'U) = 0$$

$$E (U_i, U_j) = 0 \quad i \neq j$$

وعندما لا يتحقق الفرض بان لمعادلة الانحدار اكثر من متغير توضيحي فينبغي ان تكون المتغيرات التوضيحية غير مرتبطة مع بعضها خطيا

$$\text{Var- Cov} (X'X) = 0$$

فستظهر مشكلتي التعدد الخطي و الارتباط الذاتي و التي لها تأثيرات غير جيدة على عملية التقدير وبالتالي يتوجب ان يتم الكشف عنهما وايجاد الحلول المناسبة لمعالجة هذه المشاكل والحصول على تقديرات افضل . ونظرا لكون هاتين المشكلتين لها تأثيرات سلبية على عملية التقدير فقد اهتم الكثير من الباحثين بهذه المشاكل وحاول العديد منهم وضع عدة طرائق تقدير من اجل معالجتها والتوصل الى حل للمشكلتين معا منهم G. Trenkler^[6] حيث قام بدمج مقدرات المربعات الصغرى العامة مع مقدر انحدار الحرف الاعتيادي وايضا قام الباحثين (Sakallıoglu, Siray and Kaçiranlar)^[12] في عام (2012) م بتعديل مقدر صف (r-k) باضافه مقدر المربعات الصغرى العامة GLS اليه لكي يتخلص من مشكلتي الارتباط الذاتي والتعدد الخطي في حال ظهورها معا بنفس الوقت وكما اجروا دراسة محاكاة لمقارنة اداء المقدر المقترح مع مقدر المركبات الرئيسية (PCR) (Principal Component Regression) ومقدر (ORR) (باستعمال معياري (MMSE) (Matrix mean Square Error) و (SMSE) (Scalar Mean Square Error) واشارت نتائج دراسة المحاكاة التي اجريت بان مقدر صف (r-k) الجديد هو افضل من بقية المقدرات وباستعمال معيار SMSE. وقدم الباحثان (Chandra and Tyagi)^[5] في عام (2017) م مقدر r-(k,d) class من اجل معالجة مشكلة الارتباط الذاتي حيث تفوق المقدر الجديد على نظيراته من المقدرات التي تم المقارنة معها باستعمال معيار متوسط مربعات الخطأ ودراسة خصائص المقدر المقترح .

ان الهدف الرئيسي لكل مؤسسة انتاجية هو الوصول الى اعلى نسبة وجودة للانتاج وباقل التكاليف ولهذا اصبحت قضية الانتاج هي احد الاهداف المهمة التي تسعى اليها المؤسسات الاقتصادية الكبرى فلذلك اهتم الباحثون باهم المشاكل والمعوقات التي تقف امام هذه القضية وقاموا باجراء العديد من الدراسات والبحوث لايجاد الحلول المناسبة لهذه المشاكل وبالتالي تحسين مستوى الانتاج وفي هذا البحث سنقوم بدراسة بيانات حقيقية والتي تخص انتاج مادة الزاهي المصنعة من قبل معمل المامون في قطاع الشركة العامة لصناعة الزيوت النباتية والتابعة الى وزارة الصناعة والمعادن والتي تعاني من وجود مشكلتي التعدد الخطي والارتباط الذاتي معا . ان هدف البحث هو تقدير نموذج الانحدار الخطي المتعدد لكمية انتاج مادة الزاهي والذي يعاني من مشكلتي الارتباط الذاتي والتعدد الخطي معا من خلال استعمال طريقة تقدير مدمجة لتقدير المعلومات عند وجود هاتين المشكلتين معا في نفس الوقت وهي طريقة r-(k,d) class

2. مشكلة التعدد الخطي (The Multicollinearity Problem) وطرائق الكشف عن وجودها وهي إحدى مشاكل الانحدار الخطي المتعدد وتحدث عند خرق أحد فروض التحليل ويكون هنالك ارتباط خطي بين المتغيرات التوضيحية في نموذج الانحدار فيكون هنالك صعوبة في فصل أثر كل متغير على المتغير المعتمد لهذا لا يمكن في هذه الحالة استعمال طريقة المربعات الصغرى الاعتيادية OLS. فعندما تكون محدد مصفوفة المعلومات مساويا الى الصفر $|X'X| = 0$ ستكون رتبة المصفوفة غير كاملة اي تصبح الرتبة اقل من عدد المعاملات فلا نستطيع ايجاد معكوس المصفوفة وبالتالي عدم امكانية ايجاد مقدرات للانموذج مما تؤدي الى حدوث مشكلة التعدد الخطي التام (Perfect Multicollinearity) [1]. اما في حال كون محدد مصفوفة المعلومات قريبة من الصفر $|X'X| \approx 0$ فستكون معاملاتها ذات تباينات كبيرة نتيجة زيادة في حجم الاخطاء المعيارية وتكون عملية التقدير هنا غير مجدية لكون التقديرات التي سنحصل عليها غير دقيقة مما يتسبب في ظهور مشكلة التعدد الخطي الشبه التام (Semi Perfect Multicollinearity) [1]. توجد عدة طرائق للكشف عن وجود مشكلة التعدد الخطي واهمها :

1-2- مقياس عامل تضخم التباين (Variance Inflation Factor) VIF

قام Marquarat عام 1970 باقتراح مقياس يكشف عن مشكلة التعدد الخطي في تحليل الانحدار المتعدد عندما يكون هنالك ارتباط بين المتغيرات التوضيحية في الانموذج حيث يقيس مدى تضخم التباين للمتغير التوضيحي وارتباطه بالمتغيرات التوضيحية الاخرى ويعتمد هذا المقياس على معامل التحديد ويكتب بالصيغة الآتية:

$$VIF = \frac{1}{1-R_j^2}$$

اذ ان R_j^2 : معامل الارتباط (معامل التحديد) للانحدار بين المتغير التوضيحي X_j وبقية المتغيرات التوضيحية الاخرى فاذا كانت قيمة المقياس مساوية الى الواحد فيدل ذلك على عدم وجود مشكلة تعدد خطي اما اذا كان المقياس اكبر من 5 فيدل ذلك على وجود مشكلة تعدد خطي [1]

2-2- مؤشر الحالة (Condition Index)

يقوم هذا المقياس على اساس قسمة اكبر جذر مميز على كل الجذور المميزة بصورة متتالية وحسب الصيغة الآتية:

$$C.I = \sqrt{\frac{\lambda_{\max}}{\lambda_i}} \quad i = 1, 2, \dots, P$$

حيث ان λ : تمثل الجذور المميزة لمصفوفة المعلومات

و اذا كانت قيمة اقل من 15 فهذا يدل على عدم وجود مشكلة التعدد الخطي اما اذا كانت قيمة اكبر من 15 فيدل على وجود مشكلة التعدد الخطي اما لو كانت قيمة اكبر من 30 فذلك يبين مدى خطورة مشكلة التعدد الخطي [1]

3. مشكلة الارتباط الذاتي (The Autocorrelation problem) واختبار وجودها يحدث الارتباط الذاتي نتيجة وجود ارتباطات بين الأخطاء أي أن الأخطاء غير مستقلة عن بعضها البعض عند تقدير انموذج الانحدار الخطي (1) المتعدد أي أن التباين المشترك لا يساوي صفر $E(u_i, u_j) \neq 0$ بحيث $i \neq j$ وتعتبر هذه المشكلة اختراق لأحدى فروض طريقة المربعات الصغرى الاعتيادية حيث أن المعلمات لا تكون (Best linear unbiased Estimator BLUE) (افضل تقدير خطي غير متحيز) وأن من اسباب حدوثها هو وجود مشكلة في البيانات التي تم الحصول عليها بشكل متكرر مع مرور الزمن أي البيانات مرتبطة مع نفسها وتظهر هذه المشكلة أيضا في الدراسات التي تأتي على شكل سلاسل زمنية أو عند حذف بعض المتغيرات التوضيحية من الانموذج أو نتيجة لاجراء تعديل على البيانات الموجودة أو نقوم بتقدير قيم بعض المشاهدات بالاعتماد على قيم مشاهدات أخرى فسوف تتكون علاقة بين اخطاء تلك المشاهدات وسوف تؤثر على طبيعتها وتوزيعها. توجد هناك طرائق عديدة للكشف عن مشكلة الارتباط الذاتي بين الأخطاء العشوائية للسلسلة الزمنية قيد البحث ومن أهمها :

1-3- اختبار ديربن واتسن (Durbin- Watson Test)

قدم كلا من (Durbin- Watson) عام 1950 م بايجاد اختبار يكشف عن وجود مشكلة الارتباط الذاتي والذي يعتمد على الأخطاء العشوائية في انموذج الانحدار الخطي العام كما يستعمل للارتباط الذاتي من الدرجة الاولى فقط ويعد هذا الاختبار من افضل و اكثر الاختبارات استعمالاً لسهولة تطبيقه و يقوم بحساب الأخطاء العشوائية للانموذج ولغرض اجراء هذا الاختبار بدءاً نقوم بوضع الفرضية التالية:

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

و على افتراض ان الفروق العشوائية تأخذ الصيغة الآتية :

$$U_i = \rho U_{i-1} + e_t$$

$$U_i = Y_i - \hat{Y}_i$$

ونكتب صيغة الاختبار بالشكل الآتي

$$D.W = \frac{\sum (U_i - U_{i-1})^2}{\sum U_i^2}$$

و ان قيمة (D.W) تتراوح بين الصفر والاربعة فاذا كانت (D.W) قريبة من الصفر فيدل على وجود ارتباط ذاتي موجب واما اذا كانت قيمة (D.W) تقترب من الاربعة فهذا يدل على وجود ارتباط ذاتي سالب اما اذا كانت قيمة (D.W) تساوي اثنان فيدل ذلك على عدم وجود ارتباط ذاتي. ويتم الاعتماد على جداول Durbin- Watson لبيان مدى معنوية الاختبار فيتم اما قبول فرضية العدم او رفضها. [1]

4- مقدر r(k,d) class :

لقد حظيت مشكلتي التعدد الخطي والارتباط الذاتي بأهتمام الكثير من الباحثين لكثرة ظهورهما معا وتأثيراتها السيئة على عملية التقدير فيتوجب الكشف عنها وايجاد الحلول الملائمة للتخلص منها فتوصل الكثير منهم الى ايجاد طرق جديدة عن طريق دمج مقدرين فقط

ولكنها كانت غير مجدية وخاصة اذا كان هنالك تعدد خطي تام كبير ولكن في السنوات الاخيرة توصل باحثين اخرين الى دمج اكثر من مقدر حيث وجدوا بأن الدمج سوف يحمل مزايا جميع المقدرات وبالتالي الحصول على مقدرات افضل وبالتالي الحصول على اقل متوسط مربعات خطأ ممكن (MSE) وسنستعرض في هذا البحث احد تلك الطرق وهو مقدر $r-(k,d)$ class حيث قدم الباحثان [5] (Chandra and Tyagi) في عام 2017 طريقة تقدير جديدة تقوم بمعالجة مشكلتي التعدد الخطي والارتباط الذاتي معا وتدعى بمقدر $r-(k,d)$ class وذلك عن طريق دمج عدة مقدرات مع بعضها وهي مقدر المربعات الصغرى الاعتيادية (OLS) ومقدر انحدار الحرف الاعتيادي (ORR) ومقدر انحدار المركبات الرئيسية (PCR) ومقدر ليو LIU و مقدرات المعلمتين (TP) ومقدر $r-(d)$ class ومقدر $r-(k)$ class ومقدر المربعات الصغرى العامة (GLS).

$$\hat{\beta}_r(d) = T_r(T_r'X'XT_r + I_r)^{-1}(I_r + d(T_r'X'XT_r)^{-1})T_r'X'Y \dots\dots (2)$$

$$0 < d < 1 \text{ و } r \leq p \text{ ان}$$

ان صيغة مقدر $r-(k,d)$ class تكون على النحو التالي :

$$\tilde{\beta}_r(K,d) = T_r(T_r'X'V^{-1}XT_r + kI_r)^{-1}(T_r'X'V^{-1}Y + KdT_r'\tilde{\beta}_r) \dots\dots(2)$$

$$\tilde{\beta}_r = T_r(T_r'X'V^{-1}XT_r)^{-1}T_r'X'V^{-1}Y \text{ و ان}$$

وبالتعويض عن $\tilde{\beta}_r$ فان صيغة المقدر ستكون كالآتي :

$$\tilde{\beta}_r(K,d) = T_r(T_r'X'V^{-1}XT_r + kI_r)^{-1}(T_r'X'V^{-1}XT_r)^{-1}(T_r'X'V^{-1}XT_r + KdI_r)T_r'X'V^{-1}Y \dots\dots(3)$$

كما يمكن كتابة مقدر $r-(k,d)$ class كما مبين في ادناه :

$$\begin{aligned} \tilde{\beta}_r(K,d) &= T_r S_r(K)^{-1} \Lambda_r^{-1} S_r(Kd) T_r' X' V^{-1} Y \quad (36.2) \\ &= A_1 \hat{\beta}_{GLS} \end{aligned}$$

$$A_1 = T_r S_r(K)^{-1} S_r(Kd) T_r'$$

$$S_r(q) = \Lambda_r + qI_r, q=1,k,d,kd \text{ و ان}$$

$$S_r(kd) = \Lambda_r Kd I_r$$

ان متوسط مربعات الخطا لمقدر $r-(k,d)$ class تكون كالآتي [8] :

$$MSE \tilde{\beta}_r(K,d) = \sigma^2 T_r S_r(K)^{-1} S_r(Kd) \Lambda_r^{-1} S_r(Kd) S_r(K)^{-1} T_r' + T B_1 T' \beta \beta' T B_1 T' \dots\dots(4)$$

$$B_1 = \begin{pmatrix} k(1-d)S_r(K)^{-1} & 0 \\ 0 & I_{p-r} \end{pmatrix} \text{ حيث ان}$$

وان قيمة k المثلى لمقدر $r-(k,d)$ class تعطى كالآتي :

$$K = \frac{\sigma^2}{(\alpha_i^2 - d(\frac{\sigma^2}{(\lambda_i + \alpha_i^2)}))}, \forall i = 1, 2, \dots, r$$

وان قيمه d المثلى لمقدر r -(k, d)class تكون كالآتي :

$$d_{opt} = \frac{\sum_{i=1}^r (\frac{k\alpha_i^2 - \sigma^2}{(\lambda_i + k)^2})}{\sum_{i=1}^r (\frac{k(\sigma^2 + \lambda_i \alpha_i^2)}{\lambda_i (\lambda_i + k)^2})}$$

وان لمقدر r -(k, d)class عدة حالات خاصة وهي كالآتي :

1- عندما $d=0$ يصبح مقدر $((r - k)$ Class estimator

$$= T_r (T_r' X' V^{-1} X T_r + k I_r)^{-1} T_r' X' V^{-1} Y \tilde{\beta}_r(K, 0) = \tilde{\beta}_r(K)$$

2- عندما تكون $k=1$ سنحصل على مقدر $(r-d)$ class

$$\tilde{\beta}_r(1, d) = \tilde{\beta}_r(d) = T_r (T_r' X' V^{-1} X T_r + I_r)^{-1} (I_r + d (T_r' X' V^{-1} X T_r)^{-1}) T_r' X' V^{-1} Y$$

3- عندما تكون $k=0$ يصبح مقدر المركبات الرئيسي (PCR)

$$\tilde{\beta}_r(0, d) = \tilde{\beta}_r = T_r (T_r' X' V^{-1} X T_r)^{-1} T_r' X' V^{-1} Y$$

4- عندما تكون $k=0$ و $r=p$ فيصبح مقدر مربعات صغرى عامة (GLS)

$$\tilde{\beta}_p(0, d) = \tilde{\beta}_{GLS} = (X' V^{-1} X)^{-1} X' V^{-1} Y$$

5- عندما تكون $d=0$ و $r=p$ فيصبح مقدر انحدار الحرف (ORR)

$$\tilde{\beta}_p(k, 0) = \tilde{\beta}(K) = (X' V^{-1} X + k I_p)^{-1} X' V^{-1} Y$$

6- عندما تكون $K=1$ و $r=p$ يصبح مقدر ليو (LIU)

$$\tilde{\beta}_p(1, d) = \tilde{\beta}(d) = (X' V^{-1} X + I_p)^{-1} (X' V^{-1} Y + d \tilde{\beta}_{GLS})$$

7- عندما تكون $r=p$ يصبح هنا مقدر ذات المعلمتين (TPE)

$$\tilde{\beta}_p(K, d) = \tilde{\beta}(K, d) = (X' V^{-1} X + k I_p)^{-1} (X' V^{-1} Y + K d \tilde{\beta}_{GLS})$$

5-الجانب التطبيقي

1-5 وصف للبيانات

تعد الشركة العامة لصناعة الزيوت النباتية التابعة لوزارة الصناعة والمعادن من الشركات العريقة والمهمة في العراق لكونها كانت تتميز بجودة الانتاج وتعدده والذي يضاهي المنتجات العالمية حيث كان يتم تسويق المنتجات داخل العراق وحتى خارجه وتضم هذه الشركة خمس معامل والموجودة في مختلف المحافظات والتي تنتج عدة انواع من المنظفات وتخضع هذه المنتجات جميعها لنظام الجودة (الايزو) وكانت توزع بعض من هذه المنتجات ضمن مفردات البطاقة التموينية اضافة الى تواجدها بقوة في الاسواق المحلية ولقد اخترنا في بحثنا هذه انتاج مادة الزاهي ففي عام 1981 بدء معمل المامون بانتاج هذه المادة ويعد هذا المعمل هو اول معمل في العراق يقوم بانتاج مادة المنظف السائل الزاهي وبعبوتين هي واحد لتر و اثنين لتر .وبعد احداث عام 2003 بدء انتاج هذه الشركة بالانخفاض نتيجة الازمات الاقتصادية والامنية في البلاد و ارتفاع اسعار المواد الاولية الداخلة في الانتاج اضافة الى منافسة المنتجات المستوردة لها . فاصبحت كلفة الانتاج اكبر او مساوية لكلفة البيع وتعد هذه مشكلة كبيرة بحد ذاتها .

ولقد تم جمع البيانات كما يظهر في الجدول (1) عن منتج الزاهي المصنع من قبل مصنع المامون للفترة (2006 -2017) فقط وذلك لعدم توفر البيانات للسنوات السابقة لدى الشركة وكذلك توقف انتاج ماده الزاهي لبعض الوقت بسبب احداث عام 2003 . وكانت المتغيرات المستخدمة في هذه الدراسة هي كالآتي :

Y : تمثل كمية انتاج الزاهي سنويا وحده القياس هي (طن)

سنويا وحده القياس هي (عدد) تمثل عدد العمال في المعمل : X_1

: تمثل رواتب العمال سنويا وحده القياس هي (ملايين) X_2

: تمثل مبيعات المنتج سنويا وحده القياس (ملايين) X_3

: تمثل كلفة المواد الاولية سنويا وحده القياس هي (ملايين) X_4

: تمثل عدد ايام العمل سنويا وحده القياس (عدد) X_5

: تمثل طاقة انتاج الماكينة سنويا وحده القياس هي (طن) X_6

X_7 : تمثل كمية الانتاج المخطط لها سنويا وحده القياس هي (طن)

جدول (1) يوضح العوامل المؤثرة على كمية الانتاج لمادة الزاهي لمعمل المأمون للفترة من عام 2006 ولغاية 2017

السنة	كمية الانتاج المخطط لها	طاقة انتاج الماكينة	عدد ايام العمل	تكلفة المواد الاولية	المبيعات	رواتب العمال	عدد العمال	كمية الانتاج
2006	627.200	2455.000	6937.170	752.400	241.020	205.000	2500.000	750.000
2007	313.408	2395.000	6767.630	375.600	120.620	222.000	1500.000	500.000
2008	301.000	2320.000	14555.690	361.200	115.850	229.000	1500.000	850.000
2009	612.889	2235.000	17051.880	734.400	235.860	231.000	2500.000	750.000
2010	928.947	2185.000	16670.410	1114.800	357.530	234.000	3000.000	1000.000
2011	701.987	2093.000	15968.490	841.200	269.810	235.000	2000.000	1500.000
2012	816.428	2018.000	15625.170	979.200	314.240	240.000	4000.000	1000.000
2013	824.990	1948.000	16995.560	990.000	386.680	237.000	4000.000	900.000
2014	996.814	1940.000	16811.030	1196.400	633.520	238.000	3996.000	1000.000
2015	1139.404	1024.000	7812.580	1368.100	704.350	239.000	4000.000	1000.000
2016	1183.431	904.000	6897.050	1420.800	768.990	241.000	4000.000	1000.000
2017	987.665	803.000	6126.470	1185.600	725.100	240.000	3996.000	1500.000

اختبار البيانات. 2-5

اولا : اختبار وجود مشكلة الارتباط الذاتي :

يوجد هنالك عدة اختبارات للكشف على هذه المشكلة ومن اكثر الاختبارات واسهلها تطبيقا هو اختبار Durdin-watson ويعتمد هذا الاختبار على الاخطاء العشوائية في انموذج الانحدار وقد تم استعمال برنامج spss لاختبار البيانات الحقيقية لمنتوج الزاهي والموضحة في جدول (2) فكانت قيمة (D.W) هي 3.693 ولغرض اجراء الاختبار يتم الرجوع الى جداول Durdin-watson لاجداد القيم العليا والدنيا عند حجم عينة 12 و عدد المتغيرات التوضيحية هي 7 فكانت القيمة العليا لمعامل (D.W) هي 4 والقيمة الدنيا هي 3.63 وقد بينت النتائج بوجود مشكلة ارتباط ذاتي سالب بين الاخطاء العشوائية .

$$3.63 < 3.693 < 4$$

جدول رقم (2) يبين نتائج اختبار ديرين واتسون (D.W)

Model Summary ^a					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	1.000 ^a	1.000	1.000	.472567	3.693

a. Predictors: (Constant), PlanedPro, WorkersSalary, MachinePro, DaysNo, Salling, BasicMaterial, WorkersNo
b. Dependent Variable: Product

ثانيا: اختبار الكشف عن التعدد الخطي:

تم اجراء اختبارين للكشف عن وجود مشكلة التعدد الخطي وهما معامل تضخم التباين VIF ودليل الحالة Condition index والتي تعد من اكثر الاختبارات شيوعا والمتوفرة في برنامج SPSS . وقد بينت نتائج اختبار (VIF) الى وجود مشكلة التعدد الخطي في المتغيرات التوضيحية كافة عدا المتغير السابع والذي يمثل كمية الانتاج المخطط لها سنويا حيث كانت قيمة VIF له مساوية الى 2.054 .

اما اختبار (Condition index) فتظهر نتائج الاختبار هنا بعدم وجود مشكلة التعدد الخطي عند المكون الرئيسي الاول والذي يمثل عدد العمال في المعمل و المكون الرئيسي الثاني والذي يمثل رواتب العمال والمكون الرئيسي الثالث والذي يمثل مبيعات المنتج حيث كانت قيم المؤشر اقل من 15 , اما المكون الرئيسي الرابع والذي يمثل كلفة المواد الاولية الداخلة في صناعة الزاهي والمكون الرئيسي الخامس والذي يمثل عدد ايام العمل فتشير نتائج الاختبار الى وجود مشكلة التعدد الخطي فيها حيث كانت قيمها اكبر من 15 واقل من 30 , واما المكون الرئيسي السادس والذي يمثل كمية انتاج الماكينة والمكون الرئيسي السابع والذي يمثل كمية الانتاج المخطط لها سنويا حيث تجاوزت قيمة المؤشر ال 30 مما يدل على خطورة مشكلة التعدد الخطي . والجداول ادناه يوضح التفاصيل المذكورة اعلاه :

جدول (3) والذي يمثل قيم معامل تضخم التباين

Variable	VIF
X_1	26.934
X_2	9.150
X_3	11.743
X_4	18.805
X_5	8.473
X_6	5.613
X_7	2.054

جدول (4) والذي يمثل نتائج اختبار دليل الحالة

dimension	Condition index
1	4.098
2	10.263
3	11.514
4	24.006
5	28.656
6	42.413
7	289.182

6- تحليل النتائج

نظرا لما استنتجه كلا من Chandra, S. and Tyagi, G.^[5] بأن مقدر r-(k-d) أثبت كفاءته على بقية المقدرات لذا تم تطبيق نموذج الانحدار المتعدد على بيانات دالة الانتاج لمادة الزاهي لمعمل المامون وتم تقدير معلماته باستعمال مقدر class r-(k-d) . وكانت النتائج كما يظهر في الجدول (5) الاتي :

جدول (5) يوضح قيم معاملات انتاج ماده الزاهي لمقدر r-(k,d)

الحد الثابت	X_1	X_2	X_3	X_4	X_5	X_6	X_7
-0.00182	0.00005	0.000009	0.601	-0.0025	-0.00078	-0.00001	0.0007

ومن الجدول اعلاه نجد ان عند زياده وحدة واحدة من X_1 (عدد العمال) سيؤدي الى زيادة في Y (كمية الانتاج) بمقدار 0.00005 . واما عند زيادة وحدة واحدة من X_2 (رواتب العمال) ستؤدي لزيادة في كمية الانتاج Y بمقدار 0.000009 , وان زيادة وحدة واحدة من X_3 (المبيعات) ستؤدي الى زيادة في كمية الانتاج بمقدار 0.601 , اما بالنسبة لـ X_4 (المواد الاولية) و X_5 (عدد ايام العمل) و X_6 (طاقه انتاج الماكينة) , فان كمية الانتاج سوف تتناقص عند نقصانها وبمقدار 0.0025 و 0.00078 و 0.00001 على التوالي ولكل وحده واحدة منها حيث ان كمية الانتاج تنخفض بسبب عدم توفر المواد الاولية لارتفاع اسعارها وكذلك انقطاع العمال عن العمل بسبب الاوضاع التي مر بها البلد ويتاثر الانتاج ايضا عند عطل احد الماكائن لقلة عددها وعدم توفر السيولة المالية لاصلاحها بوقت قصير .

واما في حين زيادة وحدة واحدة من X_7 (وهي كمية الانتاج التي تم التخطيط لها) سوف تؤدي الى زيادة في كمية الانتاج وبمقدار 0.0007 . وبذلك يمكن كتابة نموذج الانحدار الخطي المتعدد الذي تم حساب قيم المتغير المعتمد التقديرية من خلاله وعلى النحو التالي :

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \hat{\beta}_3 X_{i3} + \hat{\beta}_4 X_{i4} + \hat{\beta}_5 X_{i5} + \hat{\beta}_6 X_{i6} + \hat{\beta}_7 X_{i7} + U_i$$

$$\hat{Y}_i = -0.00182 + 0.00005X_{i1} + 0.000009X_{i2} + 0.601X_{i3} - 0.0025X_{i4} - X_{i5} - 0.00001X_{i6} + 0.0007X_{i7}$$

7- الاستنتاجات

تم في هذا البحث تقدير دالة الانحدار الخطي المتعدد لكمية الانتاج لمادة الزاهي في معمل المامون الذي يعاني من وجود مشكلتي الارتباط الذاتي والتعدد الخطي باستعمال المقدر r-(k-d) حيث تم في هذا المقدر تم دمج بعض المقدرات التي تجمع بين خصائص مقدر المركبات الرئيسية (PC) ومقدر ليو (LIU) ومقدر الحرف (ORR) ومقدر GLS. وأظهرت هذه الإضافات تحسناً في قيم الجذور المميزة للمركبات الرئيسية، مما ساهم في عدم حذف ما مقداره $p-r$ من المتغيرات المتسببة بالمشكلة. مما يجعل المقدر أكثر فاعلية في معالجة مشاكل التعدد الخطي والارتباط الذاتي مقارنة بالمقدرات المفردة غير الهجينة. في الجانب التطبيقي، أظهرت النتائج نقاط القوة والضعف في عملية إنتاج مادة المنظف السائل الزاهي في معمل المامون، حيث كان أكثر العوامل تأثيراً على الإنتاج هو كمية مبيعات المنتج، بينما كانت تأثيرات بقية العوامل ضعيفة.

Reference

- 1- عبد الجبار , زينب عبد الستار (2020) " مقارنة بعض طرائق التقدير لانموذج الانحدار الخطي العام بوجود مشكلتي الارتباط الذاتي والتعدد الخطي " رسالة ماجستير، كلية الادارة والاقتصاد-جامعة بغداد.
- 2.Abd Eledum, H. Y. and Alkhalifa, A. A., (2012), "Generalized Two Stages Ridge regression Estimator GTR for Multicollinearity and Autocorrelated Errors", Canadian Journal on Science and Engineering Mathematics Vol. 3 No.
- 3.Alkhamisi, M. A., (2010), "Ridge Estimation in Linear Models with Autocorrelated Errors", Communications in Statistics—Theory and Methods, Vol. 39, pp. 2630–2644.
- 4.Arowolo T. O., Adewale F. L. and Kayode, A., (2016), "A comparative study of some method of handling multicollinearity in an autocorrelated error", African Journal of Science and Technology, Vol. 13, No. 2: pp. 68 – 72.
- 5.Chandra, S. and Tyagi, G., (2017), "A general class of biased estimators in the presence of multicollinearity with autocorrelated errors", International Journal of mathematics and statistics, Vol. 19, No. 2, pp. 30-49.
- 6.G. Trenkler, (1984), "On the performance of biased estimators in the linear regression model with correlated or heteroscedastic errors", Journal of Econometrics, Vol. 25, pp. 179-190.
- 7.Husein, S. M., (2017), "Alternative estimators in logit model in the presence of multicollinearity and heteroscedasticity with a stochastic linear restricted", International Journal of Applied Mathematics and Statistics, Vol. 56, No.4, pp. 169-181.
- 8.Husein, S. M. and Abduljabar Z. A., (2020), "Comparison of the performance of some r - (k,d) class estimators with the (PCTP) estimator that used in estimating the general linear regression model in the presence of autocorrelation and multicollinearity problems at the same time", Journal of Economics and Administrative Sciences Vol. 26 No. 121 . pp. 397–414.
- 9.Husein, S. M. and Abduljabar Z. A., (2020), "Comparison of Some Estimation Methods for the Multiple Linear Regression Model with Existence of Both Auto - Correlation and Multicollinearity Problem, Journal of Al-Rafidain University College of Science, No. 46, pp. 128-153.
- 10.Kaçıranlar, S. and Sakallıoğlu, S., (2001), "Combining the Liu estimator and the principal component regression estimator", Communications in Statistics
- 11.Özkale, M. R., (2012), "Combining the unrestricted estimators into a single estimator and a simulation study on the unrestricted estimators", Journal of Statistical Computation and Simulation, Vol. 82, No. 5, pp. 653–688.
- 12.Siray, G. Ü., Kaçıranlar, S. and Sakallıoğlu, S., (2012), " r - k Class estimator in the linear regression model with correlated errors", Springer-Verlag Berlin Heidelberg, Volume 55, No. 2, pp. 393–407.

A Partial Maximum Likelihood Method to Estimate Cox Model for Censoring Data From the Right with Application

Ibrahim Khalil Hussein

إبراهيم خليل حسن

Department of Statistics,

College of Administration & Economics,

University of Baghdad, Iraq

جامعة بغداد- كلية الإدارة والاقتصاد- قسم الإحصاء-
العراق

ibrahim.khaleel2101m@coadec.uobaghdad.edu.iq

Entsar Arba Fadam

ا.د. انتصار عريبي فدعم

Department of Statistics,

College of Administration & Economics,

University of Baghdad, Iraq

جامعة بغداد- كلية الإدارة والاقتصاد- قسم الإحصاء-
العراق

entsar.arebe@coadec.uobaghdad.edu.iq

A Partial Maximum Likelihood Method to Estimate Cox Model for Censoring Data From the Right with Application

Ibrahim Khalil Hussein

إبراهيم خليل حسن

Department of Statistics,

College of Administration & Economics,

University of Baghdad, Iraq

جامعة بغداد- كلية الإدارة والاقتصاد- قسم الإحصاء- العراق

ibrahim.khaleel2101m@coadec.uobaghdad.edu.iq

Entsar Arba Fadam

ا.د. انتصار عريبي فدعم

Department of Statistics,

College of Administration & Economics,

University of Baghdad, Iraq

جامعة بغداد- كلية الإدارة والاقتصاد- قسم الإحصاء- العراق

entsar.arebe@coadec.uobaghdad.edu.iq

Abstract

In some studies, especially clinical studies, failure (death).as a result of other (external) factors, Any additional risks that may occur that compete with the case under study are called competing risks, which affect the hazard function of the Cox model. It is referred to Censoring Data From the Right, which is affected by differences in time and the phenomenon under investigation. The time of occurrence of these risks is known between two examinations, as in the Censoring Data From the Right, depending on the nature of the data for the case under study. The impact of these risks on the hazard function is estimated based on the Cox proportional hazards model, Which was estimated through the parametric part represented by the partial maximum potential method and using numerical algorithms (Newton Raphson - Down Hill) to estimate the parameters. represented by the effect of variables on the Cox hazard function , The other nonparametric part is to estimate the effect of time on the risk function through the Kaplan-Meier formula. These methods were applied to experimental data through large-scale simulations with different sample sizes from Censoring Data From the Right to size (n=30,70,100,140) As well as four initial default values for the parameters and three models of arithmetic parameters And standard deviations . After that, it was applied to real data for people with breast cancer for a random sample size of (100) for the control data from the right in order to obtain the best estimation methods. After analyzing the simulation results and real data, the results obtained showed that the Downhill (DH) algorithm is the best estimation method compared to the Newton-Raphson (NR) algorithm Depending on the statistical criterion for comparison, the average root mean square error of the model (Average Root Mean Square Error) . After estimating the competitive risk for real data based on the cumulative risk function (CIF), which explained the effect of common variables in increasing the competitive risk on the risk function of the Cox model.

Keywords: Hazard function, Censoring Data From the Right, Cox proportional hazards model, partial Maximum likelihood, Newton-Raphson algorithm, Downhill algorithm

1. المقدمة

تعد المعولية وتطبيقاتها من المواضيع المهمة التي لها تأثير واهمية كبيرة في الحياة لغرض معالجة المشاكل التي تواجه المؤسسات الصحية والخدمية والصناعية وذلك من خلال دراسة دوال البقاء (Survival Function) ودالة الخطر (Risk Function) للأجهزة والمعدات والحالات المرضية تحت المراقبة .

تستخدم اساليب المخاطر التنافسية بشكل واسع في بحوث الطب الحيوي وخاصة في مجال السرطان والتي تحتاج الى التعامل مع العديد من النتائج المحتملة في كل مكان تقريبا , على سبيل المثال قد يكون الوفاة بالسرطان هو الحدث ذات اهمية اساسية ولكن الوفاة نتيجة اسباب اخرى غير السرطان مثلا (وظائف الكبد , سكر , وظائف الكلى ,.... الخ) هي امثلة نموذجية للمخاطر التنافسية , والتي تختلف باختلاف الحالة قيد الدراسة مع الاخذ بنظر الاعتبار وقت الفشل الحقيقي الذي لا يتم ملاحظته بشكل مباشر ولكن من المعروف انه يقع في فتره بين فحصين كما في بيانات المراقبة . فقد اخذت المعولية الدور المهم والبارز في تقدير تأثير المتغيرات المشتركة (الخارجية) والوقت على دالة الخطر ودالة البقاء التي تعتمد في بياناتها على متغيرات المخاطر التنافسية و وقت اكتشاف الحالة قيد الدراسة و وقت انتهاء المراقبة , ومن ثم تقدير تأثير تلك المتغيرات في زيادة المخاطر التنافسية .

حيث تنشأ التعقيدات (Complications) عندما يكون هناك العديد من المتغيرات المشتركة (Covariates) المتغيرة بمرور الوقت المتمثلة بأسباب او انواع الفشل المؤثرة على دالة الحدث التراكمي (Cumulative Incidence Function : CIF) .

إذ تم تطبيق أنموذج كوكس للمخاطر النسبية الذي اقترح لأول مره عام 1972 من قبل العالم D.R Cox لتقدير تأثير المتغيرات المشتركة على دالة الخطر , ان اهم ما جعل هذا الأنموذج شائع الاستعمال في التحليل هو امكانية الاستفادة منه في تقدير جزء من وسطائه فقط .

ان الهدف الاساسي في تقدير معلمات اي أنموذج هو اختيار افضل طرائق تقدير تم الحصول من خلاله على مقدرات جيدة وذلك من خلال اختيار الاسلوب الامثل والصيغة الفضلى في الحصول على مقدرات المعلمات وبالاكتفاء على معيار المقارنة المستعمل في الدراسة , فهناك عدة طرائق في تقدير معلمات أنموذج كوكس شبه المعلمي لتقدير تأثير تلك المتغيرات والوقت على دالة الخطر المؤثرة على دالة البقاء للحالات قيد الدراسة من خلال تقدير الجزء المعلمي β (المتمثل بتأثير المتغيرات على دالة الخطر وذلك باستعمال طريقة الامكان الاعظم الجزئي (Partial Maximum Likelihood Method) والحصول على تقدير المعلمات من خلال الخوارزميات العددية (نيوتن رافسون – داون هيل) , اما الجزء اللامعلمي فقد تم تقديره من خلاله صيغة كابلان ماير لتقدير تأثير الوقت على دالة الخطر التي سيتم تطبيقها على بيانات المراقبة من اليمين (right-censored data) , وتقدير المخاطر التنافسية من خلال دالة الخطر التراكمي (CIF) .

1.1 الدراسات السابقة

هناك العديد من الدراسات التي ناقشت موضوع المخاطر التنافسية وتأثيرها على دالة الخطر ومنها : (Jewell et al 2003) أجرى دراسة حول تقدير توزيعات البقاء على قيد الحياة لبيانات المراقبة في سياق المخاطر التنافسية و تقدير المشاهدات الخاصة بتقدير الامكان الاعظم غير المعلمي وهو تقدير التوزيع العمري في كل من الطمث الطبيعي والجراحي . (Groeneboom et al 2008) استخدم التقدير غير المعلمي لبيانات المراقبة مع المخاطر التنافسية , حيث كان التركيز الرئيسي على مقدر الامكان الاعظم (MLE)

والمقدر البسيط حيث اثبت ان كلا المقدرين يتقاربان عالميا ومحليا بمعدل $n^{1/3}$. Tang et al (2013) طبق نموذج كوكس للمخاطر النسبية لتحليل بيانات الاعطال المبكر في كابلات الطاقة ، يقوم نموذج كوكس بتحليل مجموعة من المتغيرات المشتركة في وقت واحد وتحديد المتغيرات التي لها تأثير كبير على فشل الكابل . (Hudgens et al (2014) درس التقدير المعلمي لدالة الخطر التراكمي (CIF) لبيانات المخاطر التنافسية للمراقبة الفاصلة والذي يعتمد على تقدير معلمي مسبق ، اضافة الى مقدر الامكان البسيط والذي يستخدم جزءا فقط من البيانات المرصودة . حيث اعطى المقدر البسيط تقديرا منفصلا للنماذج لكل سبب على عكس الامكان الاعظم الكامل الذي يناسب جميع النماذج في وقت واحد . (Mao and Lin (2017) استخدم نموذج الانحدار شبه المعلمي لصياغة تأثير المتغيرات المشتركة على دالة الخطر التراكمي التي تحتاج الى تقدير ، حيث استخدم مجموعة واسعة من النماذج التحويل شبه المعلمية والتي تمتد الى النماذج (Fine) و (Gray) . (Do and Yang (2017) اقترح عدة طرق لتحليل بيانات المخاطر التنافسية في حالة فقدان اسباب الفشل ووقت الفشل ، حيث طبق منهج (Klein-Andersen's) للقيمة الزائفة ، تم تقييم الطرق المقترحة من خلال المقارنة مع تحليل الحالة الكاملة في العديد من اعدادات المحاكاة . (Thackham and Ma (2020) طبق نموذج كوكس للتعامل مع المخاطر غير المتناسبة من خلال تقدير الامكان الجزئي للمتغيرات المشتركة المتغيرة بمرور الوقت ، حيث تم معالجة الافتقار الى الدقة في تقدير معاملات الانحدار في العينات الصغيرة من خلال تطوير الامكان الاعظم لتقدير معاملات الانحدار والمخاطر الاساسية ، حيث اظهرت من خلال المحاكاة انها تتمتع بدقة متزايدة بالمقارنة مع بتقدير الامكان الجزئي (PL) في العينات الصغيرة . (Guerrero et al (2023) اجري تحليل تلوي (meta) لبيانات المريض الفردية (IPD) لتقييم تأثير منع تخثر الدم على جميع اسباب الوفاة لدى المرضى الذين يعانون من مرض تليف الكبد وتجلط الوريد البابي ، فقد اظهرت التحليلات التلوية السابقة السلامة وفعالية منع تخثر الدم في اعادة استئناء تجلط الوريد البابي لدى مرضى تليف الكبد ما اذا كانت هذه الفائدة تترجم الى تحسين البقاء على قيد الحياة غير معروف . مشكلة البحث هي حدوث مخاطر اضافية تؤثر على دالة الخطر مع الاخذ بنظر الاعتبار زمن الفشل الفعلي ، وتؤثر هذه المخاطر على زيادة دالة الخطر التراكمي (CIF) المتمثلة بالمخاطر المتنافسة على دالة الخطر الخاصة بالنموذج .

الهدف من البحث هو تقدير تأثير المتغيرات المشتركة المتمثلة بالمخاطر المتنافسة على دالة الخطر الخاصة بالنموذج لبيانات المراقبة من اليمين وتحديد الطريقة الافضل اعتمادا على المعيار الاحصائي للمقارنة وهو جذر متوسط مربع الخطأ .

2. المواد والاساليب

2.1 نموذج كوكس للمخاطر النسبية :

هو من النماذج واسعة الاستعمال في تحليل البقاء لبيانات المراقبة في التجربة وكذلك في المجالات المعتمدة على الوقت بالدرجة الاساسية في الدراسة ، فهو بديلا عن النماذج المعتاد استعمالها مثل أنموذج الانحدار الخطي (Linear Regression Model) و أنموذج الانحدار اللوجستي (Logistic Regression Model) التي لا يصلح استعمالها مع بيانات المراقبة ، يعد هذا الأنموذج من النماذج المشهورة التي تدرس بيانات المراقبة وذلك لسهولة التعامل مع البيانات (Tang et al, 2013) . يعد أنموذج كوكس أنموذج شبه معلمي لكونه يحتوي على جزئين الجزء الاول معلمي والذي يمثل الدالة الاسية (Exponential Function) والجزء الاخر اللامعلمي والمتمثل بتأثير الوقت على دالة الخطر الاساسية . (Risk Function) حيث يمكن الاستفادة من أنموذج كوكس بتقدير جزء فقط من وسطائه وهذا ما جعله واسع الاستعمال ، وان الصيغة الرياضية لهذا الأنموذج (Thackham et al, 2020) :

$$h(t, x) = h_0(t).exp(\beta^T X) \quad (1)$$

حيث:

$h_o(t)$: والتي تسمى بدالة المخاطرة الاساسية Baseline Hazard Function ، β هو متجه معاملات أنموذج Cox .

$X_i = (X_1, X_2, \dots, X_n)$ هي مصفوفة المتغيرات المستقلة المتوقع تأثيرها على دالة المخاطرة $h(t, x)$.

إذا كان لدينا متغيرين مستقلين من X_1 و X_2 ، فإن نسبة الخطر للمتغير الاول X_1 الى المتغير الثاني X_2 يتم حسابها كالآتي (Scheike and Sun, 2007).

$$\frac{h(t, x_1)}{h(t, x_2)} = \frac{h_o(t) \cdot \exp(\beta x_1)}{h_o(t) \cdot \exp(\beta x_2)}$$

$$\frac{h(t, x_1)}{h(t, x_2)} = \exp[\beta(x_1 - x_2)]$$

وبعد اخذ اللوغاريتم الطبيعي يصبح :

$$\ln \left[\frac{h(t, x_1)}{h(t, x_2)} \right] = \beta(x_1 - x_2)$$

هذا يعني ان β تمثل الزيادة في اللوغاريتم الطبيعي لنسبة خطر المتغير الى نسبة خطر المتغير الثاني بمقدار وحدة واحدة $\frac{h(t, x_1)}{h(t, x_2)}$.

2.1.1 طرق تقدير نموذج كوكس:

2.1.1.1 طريقة الامكان الاعظم الجزئي:

اقترح العالم Cox عام (1980) طريقة لتقدير معلمات β والتي تسمى طريقة الامكان الجزئي (Partial Likelihood Method) فهي طريقة واسعة الاستعمال والتي تعتمد على ترتيب الاحداث (الموت) بالشكل الآتي : $t_{(1)}, t_{(2)}, \dots, t_{(r)}$ ، وان مجموع العناصر تحت

الخطر (الموت) عند t هو $F(t_i)$ ، في البداية تفترض هذه الطريقة في كل حدث $t_{(i)}$ يوجد عنصر واحد حدث له الحدث المدروس، تم استعمال طريقة الامكان الجزئي (PL) في تقدير معالمه والتي تتميز انها لا تعتمد على $h_o(t)$ كما موضح في الصيغة الآتية (

. Cox, 1975).

$$f(\beta) = \frac{h(t_{(i)}, x_k)}{\sum_{j \in F(t_{(i)})} h(t_{(i)}, x_j)} \quad (2)$$

حيث $h(t(i), x_k)$ تمثل دالة الحدث (Event function) المدروس في اللحظة الزمنية $t(i)$ و $h(t(i), x_j)$ تمثل دالة وقوع حدث الاهتمام (الموت) على كل عنصر من العناصر تحت الخطر $F(t(i))$.

بعد ذلك فإن الامكان الجزئي يعمل على افتراض الحدث محل الاهتمام (الوفاة) في وقت ما $t(i)$ ، والتي يمكن كتابتها على النحو التالي (Sinha et al, 2003):

$$L(\beta) = \prod_{i=1}^r \frac{\exp(\beta^T x_k)}{\sum_{j \in F(t(i))} \exp(\beta^T x_j)} \quad (3)$$

وبأخذ اللوغاريتم لدالة الامكان الجزئي :

$$\ln l(\beta) = \sum_{i=1}^r \{x_i \beta^T - \ln[\sum_{j \in F} \exp(x_j \beta^T)]\} \quad (4)$$

يمكن الحصول على تقدير المعلومات من الامكان الجزئي باستعمال الطرق العددية ، ومن الاساليب العددية الشائع استعمالها هي خوارزمية نيوتن رافسون (NR) وخوارزمية داون هيل (DH) للدالة (4) .

(a) خوارزمية نيوتن رافسون:

وهي خوارزمية فعالة لإيجاد الجذور لدالة القيم الحقيقية او الحصول على افضل تقريب للأصفر اي تستعمل لإيجاد الحدود الدنيا والحدود العليا للمعلومات من خلال إيجاد جذور المشتقة الاولى للمعلمة . وتستعمل أيضاً لحل المعادلات الخطية وإيجاد حل لجذور المعادلات غير الخطية التي تعد معادلات معقدة لا يمكن حلها بالطرائق التقليدية ، التي تم اقتراحها من قبل (اسحاق نيوتن وجوزيف رافسون) في عام (1960) ويرمز لها بالرمز (NR) (Yalçinkaya et al, 2018) .

أن الهدف الذي دفعنا لاستعمال هذه الخوارزمية هو الحصول على أفضل تقديرات للمعلومات (β) بالاعتماد على التقديرات الاولى (يتم الحصول على التقديرات الاولى من طريقة الامكان الجزئية) . وباستخدام الاسلوب التكراري لخوارزمية نيوتن رافسون لمعادلة الامكان الجزئي رقم (4) المشتقة الاولى للوغاريتم دالة الامكان الجزئي بالنسبة للمعلومات ، وسيتم توضيح عمل الخوارزمية بالتفصيل، وتأتي خوارزمية (NR) وفقاً للصيغة الرياضية الاتية (Akram and Ann , 2015) :

$$\hat{\beta}^{(r+1)} = \hat{\beta}^{(r)} - \frac{F^{(r)}}{P^{(r)}} \quad (5)$$

في سياق خوارزمية نيوتن رافسون

- r : يمثل عدد التكرارات.
- $\hat{\beta}^{(r+1)}$: يمثل تقدير المعلومات الجديدة للتكرار اللاحق $(r+1)$.
- $\hat{\beta}^{(r)}$: يمثل تقدير المعلمة التي تم الحصول عليها في r .
- $F^{(r)}$ يمثل المتجه المشتق الاول للوغاريتم لدالة الامكان الجزئي في التكرار r . (Casella and Bachmann, 2021).

$$F = \left(\frac{\partial \ln L(\beta)}{\partial \beta} \right) = \sum_{i=1}^r x_i - \frac{\sum_{j \in F} x_j e^{x_j \beta}}{\sum_{j \in F} e^{x_j \beta}}$$

$P^{(r)}$ تمثل مصفوفة المشتقة الثانية للوغاريتم دالة الامكان الاعظم للتكرار r .

$$P = \left[\frac{\partial^2 \ln L(\beta)}{\partial^2 \beta} \right] = - \sum_{j \in F} \left[\frac{(\sum_{j \in F} e^{x_j \beta})(\sum_{j \in F} x_j^2 e^{x_j \beta}) - (\sum_{j \in F} x_j e^{x_j \beta})^2}{(\sum_{j \in F} e^{x_j \beta})^2} \right]$$

بعد حساب الفرق المطلق بين تقدير المعلمات الجديدة وتقدير المعلمات السابقة، نتحقق الخوارزمية اذا كان هذا الفرق المطلق اصغر من قيمة c ونقوم بطباعة الحل الامثل نتوقف الخوارزمية عن العمل. اما في حال كانت قيمة الفرق المطلق اكبر من (c) تستمر اعادة التكرار الى ان يتحقق شرط التوقف.

$$|\beta^{(r+1)} - \beta^{(r)}| < c \quad (6)$$

حيث c هي قيمة صغيره جدا وثابتة.

(b) خوارزمية داون هيل:

تعد هذه الخوارزمية من الخوارزمية العددية التي اقترحت من قبل العالمين (Nelder and Mead) عام (1965). والتي تعمل على ايجاد الحل الامثل لدالة الهدف اي الحصول على الحد الادنى لدالة الهدف في الدوال المعقدة التي تحتاج الى تحسينات، تستعمل هذه الخوارزمية في العديد من المجالات لكونها خوارزمية سهلة الفهم والاستعمال ولعدم حاجتها الى مشتقات فهي تحظى بشعبية كبيرة في العديد من مجالات العلوم والتكنولوجيا وخاصة الكيمياء والطب والتي تعتمد على تخمين عدد من النقاط لدالة الهدف (Fajfar et al, 2017). تعتمد خوارزمية داون هيل على الشكل الهندسي (Simplex) الموجود بعدة اشكال هندسية لـ n من الابعاد و $n+1$ من النقاط والمتمثلة بـ $(Z_1, Z_2, \dots, Z_{n+1})$ والتي تمثل ترتيب قيم دالة الهدف عند كل اختبار بالشكل الاتي. (Galántai, 2021)

$$f_{z_1} \leq f_{z_2} \leq \dots \leq f_{z_{n+1}} \quad (7)$$

حيث Z_1 تمثل افضل نقطة، و Z_{n+1} تمثل اسوء نقطة.

فيما يلي خطوات تقدير المعلمة:

1. اختيار دالة الامكان الجزئي المتمثلة بدالة الهدف وهو سالب لوغاريتم الدالة. (4)

$$f_z = -\ln L(z) \quad , \quad \text{where } Z = (\beta)$$

2. تثبيت قيم معلمات الخوارزمية (الانعكاس $\alpha = 1$), (التوسع $\varepsilon = 2$), (الانكماش $\gamma = 0.5$), (التقليص $\psi = -0.5$).

3. انشاء مصفوفة S كحل اولي وبأبعاد $(n+1)*1$.

$$S = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{n+1} \end{bmatrix} \quad (8)$$

4. تقدير دالة الهدف في كل صف من صفوف المصفوفة S ورتب التقديرات من الأدنى الى الأعلى .

$$f_{z_1} \leq f_{z_2} \leq \dots \leq f_{z_{n+1}}$$

5. حساب نقطة الانعكاس (r) باستعمال الصيغة الآتية:

$$Z_r = \bar{S} + \alpha(\bar{S} - Z_{n+1}) \quad (9)$$

بعد حساب نقطة الانعكاس , نقوم بحساب دالة الهدف (f_{z_r}) اذا:

$f_{z_1} < f_{z_r} < \dots < f_{z_n}$ نجعل $Z_r = Z_{n+1}$ ثم ننتقل للخطوة (9) خلاف ذلك ننتقل للخطوة اللاحقة.

6. حساب نقطة التوسع (e) وفق الصيغة الآتية:

$$Z_e = \bar{S} + \varepsilon(Z_r - \bar{S}) \quad (10)$$

وبعد حساب التوسع نحسب دالة الهدف (f_{z_e}) اذا:

$f_{z_e} < f_{z_r}$ نجعل $Z_e = Z_{n+1}$ وننتقل الى الخطوة (9) خلاف ذلك ننتقل للخطوة اللاحقة .

7. حساب نقطة الانكماش (c) وفق الصيغة الآتية:

$$Z_c = \bar{S} + \gamma(Z_{n+1} - \bar{S}) \quad (11)$$

بعد حساب الانكماش نحسب دالة الهدف f_{z_c} اذا كانت :

$f_{z_c} < f_{z_n}$ نجعل $Z_c = Z_{n+1}$ وننتقل للخطوة 9. خلاف ذلك ننتقل للخطوة اللاحقة.

8. حساب نقطة التقليل (sh) وفق الصيغة الآتية .

$$Z_{sh} = Z_1 + \psi(Z_i - Z_1) \quad (12)$$

9. نطبق هذه الخطوة عند تحقق شرط التوقف في الخطوات السابقة وفق الصيغة الآتية.

$$\left| \frac{\max(f) - \min(f)}{\max(f)} \right| < \varepsilon$$

في حالة تحقق شرط التوقف تستمر الخوارزمية في طباعة الحل ، خلاف ذلك تعود الى الخطوة (5) وتستمر هذه العملية التكرارية حتى يتحقق التقارب المنشود. (Mehta, 2020)

2.1.1.2 كابلان ماير:

يعد مقدر كابلان ماير أو مقدر حد المنتج (Product-limit) مقدر غير معلمي يستخدم لتقدير دالة البقاء (Survival Function) على قيد الحياة ودالة الخطر، يتميز هذا المقدر بسهولة حسابه وفهمه، ففي البحث الطبي غالباً ما يستعمل لقياس نسبة المرضى الذين يعيشون لفترة معينة من الوقت بعد العلاج. وفي مجالات أخرى يمكن استعماله لتقدير طول الفترة الزمنية التي يبقى فيها الأشخاص عاطلين عن العمل بعد فقدان الوظيفة، حيث تم تسمية هذا المقدر على اسم ادوارد ل. كابلان و بول ماير عام (1958) (Smith and Smith, 2003). عادة ما يطلق على التمثيل المرئي لهذه الدالة بمنحنى كابلان ماير (Kaplan Maeier curve) فهو يوضح احتمال البقاء على محور y والوقت على محور x . حيث يفترض كابلان ماير حدث الاهتمام يحدث في وقت محدد بوضوح، وأن احتمالية البقاء لجميع المشاهدات هو نفسه ولا يهم متى دخلوا الدراسة بالضبط. المشاهدات الخاضعة للمراقبة لها نفس احتمالية البقاء على قيد الحياة، فإن دالة البقاء لـ (KM) يعرف بالصيغة الآتية Andrade, (2023):

$$\hat{S}(t) = \prod_{i=1}^k \frac{n_i - d_i}{n_i}, \quad k = 1, 2, \dots, r \quad (13)$$

حيث أن:

n_i : عدد الأشخاص الذين نجوا من الوفاة حتى الوقت t_i .

d_i : عدد الاحداث (الوفيات) التي حدثت في الوقت t_i .

إن دالة التوزيع الفشل التراكمي تعطى بالصيغة الآتية:

$$\hat{F}(t) = 1 - \hat{S}(t)$$

يتم تقدير دالة الكثافة الاحتمالية من خلال الطريقة التجريبية (Empirical Methods) غير المعلمية وفق الصيغة التالية:

$$\begin{aligned} \hat{f}(t) &= - \frac{\hat{S}(t_{i+1}) - \hat{S}(t_i)}{t_{i+1} - t_i} \\ &= \frac{1}{(t_{i+1} - t_i)(n+1)} \quad \text{for } t_i < t < t_{i+1} \end{aligned} \quad (14)$$

وإن دالة الخطر لوقت الفشل تعطى بالصيغة الآتية (Teoh, 2008)

$$\hat{h}(t) = \frac{\hat{f}(t)}{\hat{S}(t)} = \frac{1}{(t_{i+1} - t_i)(n+1 - d_i)} \quad \text{for } t_i < t < t_{i+1} \quad (15)$$

2.2 المخاطر التنافسية:

تنشأ المخاطر التنافسية عندما تكون أكثر من نتيجة في البحث السريري أثناء متابعة البقاء على قيد الحياة، على سبيل المثال قد تكون الوفاة المرتبطة بالسرطان ذات أهمية أساسية ولكن الوفاة نتيجة أسباب أخرى غير السرطان هي أمثلة نموذجية للمخاطر التنافسية، إذ يمكن الموت قبل الإصابة بالمرض (مراقبة مستقلة) حيث لا يمكن ربط تأثير المتغيرات على دالة الخطر بدالة الخطر التراكمي (CIF). لذلك سنطبق الطرائق المختلفة لتحليل البقاء في وجود مخاطر تنافسية، يمكن وصف بيانات البقاء من خلال دالة الخطر $[h(t)]$ والتي تصف المخاطر اللحظية للفشل في ظل البقاء على قيد الحياة عند الزمن t (Sildnes, 2015)، إن دالة الخطر التراكمي $[H(t)]$ هي $[h(t)]$ المضافة بمرور الوقت من 0 إلى t ، وإن دالة البقاء $\hat{S}(t)$ يمكن تقديرها لا معلمي باستعمال كابلان ماير $\hat{H}(t)$ يمكن تقديرها باستعمال مقدر نيلسون آلين (Nelson–Aalen Estimator) من خلال الصيغة الآتية (Groeneboom et al, 2008):

$$\hat{H}(t) = \sum \frac{d_i}{r_i} \quad (16)$$

حيث d_i هو عدد الأفراد الذين يموتون خلال الفترة t_i و r_i هو عدد الأفراد المعرضين للخطر خلال t_i .

3. مناقشة النتائج

3.1 المحاكاة

المحاكاة (Simulation) هي عملية تمثيل أو تقليد للواقع الحقيقي أي الحصول على صورته تشابه أو تماثل أي نموذج أو نظام من غير الاعتماد على نفس النموذج أو النظام ، غالبا ما نواجه في الواقع الحقيقي عمليات معقدة يصعب تحليلها تحليلًا منطقيًا مما يؤدي إلى ترجمة النظريات الاحصائية من خلال المحاكاة للحصول على نتائج (مقدرات) تحاكي الواقع الحقيقي ، يتم أخذ عينات بأحجام مختلفة (صغيرة ، متوسطة ، كبيرة) للوصول إلى الحل الأمثل أي الحصول على نتائج مشابهة للصور الحقيقية والتي تحقق لنا رؤية بصرية للعملية قيد الدراسة أو الواقع الحقيقي من خلال المحاكاة . حيث تم الاعتماد على طريقة مونت كارلو في توليد بيانات المحاكاة تستعمل المحاكاة للمقارنة بين طرائق التقدير المدروسة لمعرفة الطريقة الفضلى للتقدير ، هذا ما انصب عليه اهتمامنا للمقارنة بين طريقتي الأماكن الأعظم الجزئي مضروب بدالة الخطر الأساسية المقدرة بصيغة كابلان ماير تلقائيا لكل تجربة لبيانات المراقبة من اليمين وتحديد أي الطريقة الفضلى في التقدير (Thackham and Ma, 2020).

3.1.1 خطوات المحاكاة:

المرحلة 1:

تم توليد قيم المتغيرات المشتركة (X_1, X_2, X_3) المتمثلة بالمخاطر التنافسية لبيانات المراقبة من اليمين عشوائيا اعتمادا على التوزيع المنتظم (Uniform distribution) للفترة (0,1) للقيمة الأولية ، بعد ذلك تولد البيانات من خلال التوزيع الطبيعي (Normal Distribution) اعتمادا على الأوساط الحسابية والانحرافات المعيارية التي تم افتراضها. والمعرفة وفق الصيغة الآتية ::

$$f(x) = \frac{x - \mu}{\sigma}$$

حيث ان :

μ : يمثل الوسط الحسابي المفترض ، σ يمثل الانحراف المعياري المفترض .

اما بالنسبة للزمن t فقد تم توليده بشكل متجه (Vector). $t=1,2,...,n$.

المرحلة 2:

تعتبر هذه المرحلة من اهم المراحل التي يتم فيها تحديد القيم الأولية المفترضة ، حيث تم تحديد القيم على النحو التالي :

- افتراض حجم العينة المطلوبة : $n=(30,70,100,140)$.
- تحديد التكرار لكل وحدة تجريبية 1000 .
- تحديد قيم المعلمات الأولية β لمقدر نيوتن رافسون : $(0.242, 1.342, 0.842)$ ولمقدر داون هيل : $(1.1778, 0.7661, 2.1023)$. يتم عرض المتوسطات الحسابية والتباينات في الجدول التالي :

الجدول 1: يعرض الوسط الحسابي والانحراف المعياري للمتغيرات المشتركة

Model	Mean(t)	$\mu(X_1)$	$\mu(X_2)$	$\mu(X_3)$	Std(t)	$\sigma(X_1)$	$\sigma(X_2)$	$\sigma(X_3)$
I	1584.35	5.85	23.03	4.83	318.36	1.87	16.57	1.68
II	1742.79	6.44	25.33	5.31	350.20	2.05	18.23	1.85
III	1425.92	5.27	20.73	4.34	286.53	1.68	14.91	1.51

حيث $\mu(t)$: يمثل متوسطات حسابية للوقت قريبا من المتوسط الزمني للبيانات الحقيقية .

المرحلة 3:

تقدير المعلمات باستعمال الخوارزميات العددية (نيوتن رافسون وداون هيل) لطريقة الامكان الجزئي للجزء المعلمي . اما بالنسبة للجزء الغير معلمي فقد تم تقديره تلقائيا لأحجام عينات مختلفة باستعمال كابلان ماير . وذلك لغرض الحصول على دالة الخطر لنموذج كوكس كما هو موضح في المعادلة (1) وحساب المخاطر التنافسية ودالة البقاء كما في المعادلة (16) و (13) على التوالي.

المرحلة 4:

تمت المقارنة بين طرق التقدير المدروسة وتحديد افضلها اعتمادا على المعيار الاحصائي (ARMSE) لنموذج كوكس حسب الصيغة التالية :

$$RMSE = \sqrt{MSE h_{(t,x)}} \quad (17)$$

حيث :

$$AMSE h_{(t,x)} = \frac{\sum_{i=1}^r (\hat{h}_t - \hat{h}_t)^2}{r} \quad (18)$$

حيث :

- r : يمثل عدد التكرار لكل تجربة .
- \hat{h}_t : يمثل مقدر دالة الخطر للقيم المقترضة .
- $\hat{\hat{h}}_t$: تمثل دالة الخطر المقدرة لنموذج كوكس .

3.1.2 نتائج تجربة المحاكاة :

استنادا للنتائج التي تم الحصول عليها من الطرق المختلفة واحجام العينات المختلفة ونماذج مختلفة من المعلمات ، سيتم اجراء المقارنة لتحديد افضل طريقة تقدير اعتمادا على معدل جذر متوسط مربع الخطأ (ARMSE) .

الحالة الاولى : تمثل هذه الحالة المعلمات المفترضة $\hat{\beta}(NR) = (0.242, 1.342, 0.842)$ مع النماذج الثلاثة للأوساط الحسابية والانحرافات المعيارية والمعرفة بالرموز (I) (II) (III) ولأحجام العينات المختلفة المبينة في الجدول 2 .

الجدول : 2 يوضح قيم (ARMSE) لدالة الخطر لأنموذج Cox للمعلّمتات و للنماذج الثلاثة ولأحجام عينات مختلفة للمقارنة بين طرائق التقدير بالنسبة لبيانات المراقبة من اليمين

Models of the initial parameters	N	Methods	ARMSE(β_1)	ARMSE(β_2)	ARMSE(β_3)	ARMSEh _(t,x)	Best Method
I	30	NR	0.6958	0.7252	0.4550	0.3567	DH
		DH	0.9127	0.9226	0.9088	0.0033	
	70	NR	0.7229	0.6982	0.4510	0.2436	DH
		DH	0.9235	0.9172	0.9237	0.0026	
	100	NR	0.7289	0.6923	0.4505	0.1634	DH
		DH	0.9234	0.9235	0.9233	0.0021	
	140	NR	0.7258	0.6953	0.4507	0.1451	DH
		DH	0.9249	0.9251	0.9252	0.0016	
II	30	NR	0.7422	0.6777	0.4497	0.4816	DH
		DH	0.9245	0.9212	0.9105	0.0013	
	70	NR	0.7072	0.7137	0.4529	0.1236	DH
		DH	0.8988	0.9010	0.9031	0.0013	
	100	NR	0.7133	0.7076	0.4521	0.0934	DH
		DH	0.9230	0.9252	0.9246	0.0015	
	140	NR	0.7173	0.7036	0.4516	0.8874	DH
		DH	0.9185	0.9200	0.9179	0.0008	
III	30	NR	0.7142	0.7067	0.4520	0.5488	DH
		DH	0.9128	0.9231	0.9238	0.0044	
	70	NR	0.7032	0.7177	0.4536	0.6628	DH
		DH	0.9179	0.9168	0.9236	0.0029	
	100	NR	0.7301	0.6911	0.4509	0.0739	DH
		DH	0.9244	0.9241	0.9245	0.0016	
	140	NR	0.7201	0.7009	0.4513	0.5869	DH
		DH	0.9221	0.9248	0.9196	0.0027	

في الجدول (2) نلاحظ كلما زاد حجم العينة كلما انخفض قيمة المعيار الاحصائي للمقارنة بالنسبة للمعاملات وكذلك بالنسبة لدالة الخطر لنموذج كوكس ، وهذا ما يتوافق مع النظرية الاحصائية . حيث اظهرت النتائج في الجدول (2) ان خوارزمية داون هيل (DH) اثبتت افضليتها في التقدير من خوارزمية نيوتن رافسون (NR) لمختلف احجام العينات ولمختلف النماذج .

الحالة الثانية : تمثل المعلمات المفترضة $(\beta) = (1.1778 ; 0.7661 ; 2.1023)$ مع النماذج الثلاثة للأوساط الحسابية والانحرافات المعيارية والمعروفة بالرموز (I) (II)(III) . ولأحجام العينات المختلفة الموضحة في الجدول 3 .

الجدول : 3 يوضح قيم (ARMSE) لدالة الخطر لأنموذج Cox للمعاملات و للنماذج الثلاثة ولأحجام عينات مختلفة للمقارنة بين طرائق التقدير بالنسبة لبيانات المراقبة من اليمين							
Models of the initial parameters	N	Methods	ARMSE(β_1)	ARMSE(β_2)	ARMSE(β_3)	ARMSEh _(t,x)	Best Method
I	30	NR	0.5762	0.7858	0.9624	0.6830	DH
		DH	1.4598	1.4077	1.4379	0.0024	
	70	NR	0.5809	0.7985	0.9479	4.9106	DH
		DH	1.4432	1.4313	1.4576	0.0034	
	100	NR	0.5819	0.8011	0.9450	0.2865	DH
		DH	1.4597	1.4521	1.4543	0.0023	
	140	NR	0.5817	0.8008	0.9454	0.2623	DH
		DH	1.4341	1.4335	1.4587	0.0005	
II	30	NR	0.5610	0.7249	1.0377	0.3587	DH
		DH	1.4574	1.4598	1.4524	0.0012	
	70	NR	0.5798	0.7958	0.9510	2.8378	DH
		DH	1.4385	1.4519	1.4168	0.0043	
	100	NR	0.5813	0.7996	0.9467	0.4748	DH
		DH	1.4573	1.4580	1.4598	0.0024	
	140	NR	0.5813	0.7997	0.9465	0.5893	DH
		DH	1.4465	1.4545	1.4571	0.0020	
III	30	NR	0.5706	0.7683	0.9831	0.4785	DH
		DH	1.4595	1.4598	1.4597	0.0039	
	70	NR	0.5767	0.7872	0.9609	1.1188	DH
		DH	1.4563	1.4493	1.4530	0.0023	
	100	NR	0.5812	0.7993	0.9470	0.2844	DH
		DH	1.4315	1.4485	1.4425	0.0010	
	140	NR	0.5823	0.8022	0.9438	0.2137	DH
		DH	1.4503	1.4532	1.4579	0.0013	

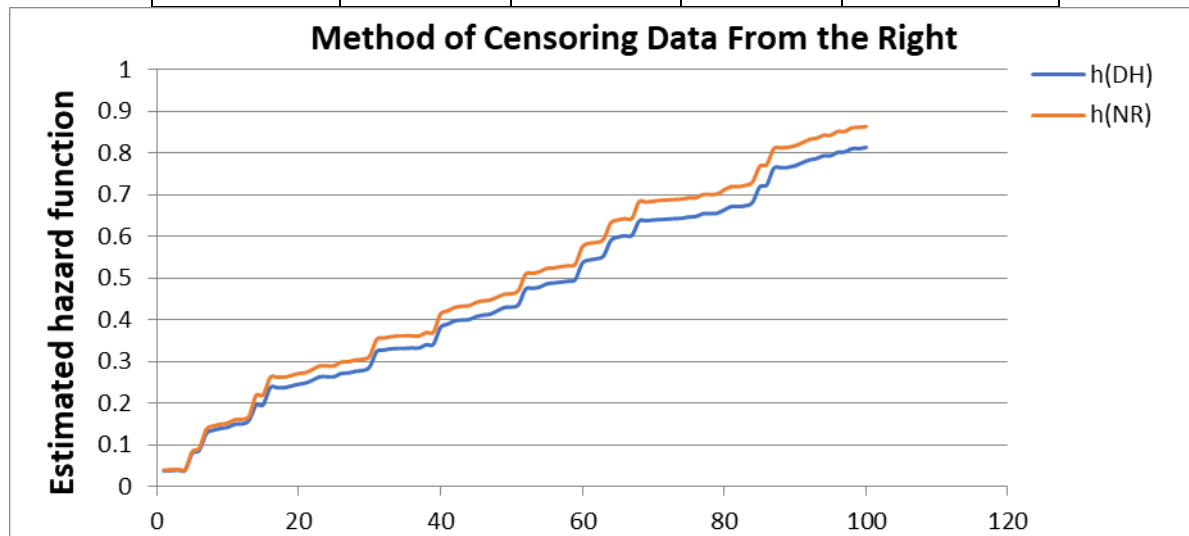
في الجدول (3) الموضح لقيم (ARMSE) للنماذج المذكورة اعلاه نلاحظ كلما زاد حجم العينة كلما كانت قيمة المعيار الاحصائي للمقارنة (ARMSE) بالنسبة لتقدير المعلمات وكذلك لدالة الخطر لنموذج كوكس وهذا ما يتوافق مع النظرية الاحصائية ، من خلال النتائج في الجدول اعلاه نلاحظ ان خوارزمية داون هيل (DH) هي الافضل من خوارزمية نيوتن رافسون (NR) من حيث دقة وكفاءة التقدير .

3.2 تطبيق البيانات الحقيقية :

تم جمع البيانات من وزارة الصحة العراقية / مستشفى الامل للأورام السرطانية لعينة حجمها 100 فردا مصاب بسرطان الثدي . ولمدة زمنية من 2019 ولغاية 2023 لمراقبة تأثير المتغيرات المشتركة (مخاطر تنافسية) ، وخلال الفترة المحددة للمراقبة تم ملاحظة حدوث الوفاة لـ (80) مصاب قبل انتهاء وقت المراقبة المحدد وهناك (20) مصاب مستمر بالبقاء على قيد الحياة ، حيث تم الاعتماد على ثلاث متغيرات بعد استشارة الاطباء المعنيين المتمثلة بـ (X_1 : وضائف الكلى ، X_2 : وضائف الكبد ، X_3 : السكر) التي تمثل المخاطر التنافسية . تم تحديد القيم الاولى للمعلمات $(\beta_1 = 0, \beta_2 = 1.5, \beta_3 = 1, \beta_4 = 4)$ للحصول على المقدّر الأولي لدالة الخطر لنموذج كوكس ، بعد ذلك تم تطبيق طرق التقدير للحصول على مقدرات المعلمة ودالة الخطر الأساسية لتأثير الوقت ثم تحديد الطريقة الأفضل في التقدير اعتمادا على المعيار الاحصائي (ARMSE) لنموذج كوكس بالصيغة (1) كما موضح في الجدول ادناه :

الجدول 4 : يوضح المعلمات المقدرة و (RMSE) لدالة الخطر لنموذج كوكس

Methods	Beta estimated			$RMSEh_{(t,x)}$
	β_1	β_2	β_3	
Newton Raphson	0.1035	1.2035	0.7035	1.0165e-05
Downhill	1.1666	0.1272	1.3130	2.3013e-07



الشكل 1 : يوضح دالة الخطر للمقدرات ولدالة الخطر لنموذج كوكس بالنسبة لبيانات المراقبة من اليمين

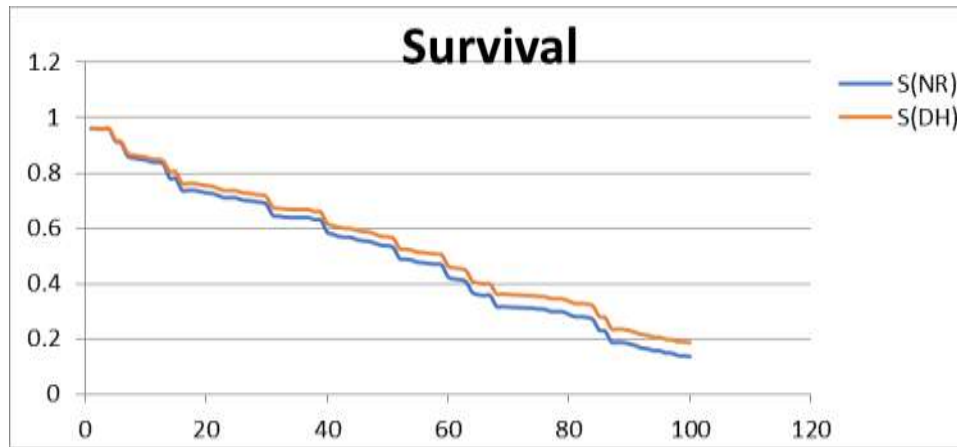
في الجدول (4) والشكل (1) الموضح لـ $(RMSEh_{(t,x)})$ لطرائق التقدير لأنموذج Cox نلاحظ ان خوارزمية داون هيل (DH) هي الطريقة الافضل في التقدير من خوارزمية نيوتن رافسون (NR) .

وبعد ذلك تم تقدير دالة البقاء $\hat{S}(t)$ لجميع الاشخاص المصابين والوقت المحدد بالأيام , لبيانات المراقبة من اليمين حيث يوضح الجدول (5) والشكل (2) تقدير البقاء لجميع الطرائق المستعملة , ومن خلال المقدرات والشكل نلاحظ تحقيق سلوك او اداء دالة البقاء كونها متناقصة مع زيادة دالة الخطر بمرور وقت الدراسة والذي يتطابق مع المعولية

الجدول 5 : يوضح تقدير دالة البقاء $\hat{S}(t)$ لبيانات المراقبة من اليمين

t	S(NR)	S(DH)	t	S(NR)	S(DH)
1	0.961313	0.963442	51	0.529782	0.563727
2	0.960233	0.962406	52	0.491462	0.527376
3	0.95907	0.961369	53	0.489361	0.525476
4	0.95863	0.960945	54	0.485336	0.521558
5	0.916612	0.920747	55	0.477428	0.513954
6	0.908032	0.912733	56	0.475708	0.51238
7	0.863144	0.872918	57	0.472938	0.509773
8	0.854936	0.865245	58	0.470928	0.507899
9	0.850623	0.861285	59	0.466823	0.503833
10	0.847743	0.85841	60	0.425534	0.464878
11	0.840041	0.850783	61	0.417336	0.457099
12	0.839326	0.850134	62	0.414657	0.454377
13	0.831756	0.842447	63	0.406771	0.446938
14	0.782045	0.805316	64	0.369065	0.41055
15	0.780896	0.804244	65	0.360551	0.402147
16	0.738952	0.763407	66	0.357804	0.39878
17	0.738047	0.762633	67	0.357045	0.398051
18	0.737936	0.762526	68	0.31845	0.363704
19	0.73357	0.758816	69	0.317615	0.362863
20	0.729299	0.75504	70	0.315571	0.360854
21	0.726613	0.752355	71	0.314319	0.359694
22	0.719127	0.744926	72	0.31343	0.358879
23	0.71111	0.736998	73	0.312273	0.357689
24	0.710891	0.73677	74	0.311382	0.356846
25	0.7107	0.736547	75	0.308329	0.354213
26	0.702387	0.728912	76	0.307619	0.353497

27	0.701217	0.727804	77	0.299694	0.34626
28	0.697223	0.723925	78	0.299652	0.346218
29	0.69457	0.721559	79	0.298525	0.345149
30	0.68751	0.714497	80	0.289053	0.336662
31	0.648166	0.677276	81	0.28082	0.328983
32	0.644114	0.673032	82	0.280562	0.328718
33	0.641391	0.670359	83	0.278474	0.326665
34	0.639423	0.6686	84	0.269765	0.318726
35	0.639376	0.668556	85	0.232562	0.281714
36	0.639289	0.668474	86	0.228133	0.277487
37	0.639151	0.668346	87	0.189957	0.237493
38	0.631484	0.660485	88	0.188494	0.236325
39	0.629812	0.658951	89	0.18698	0.234945
40	0.587226	0.618928	90	0.182744	0.230694
41	0.579082	0.610882	91	0.17583	0.223868
42	0.57088	0.603411	92	0.167523	0.216666
43	0.568182	0.600619	93	0.164909	0.213883
44	0.567146	0.599728	94	0.158031	0.206802
45	0.558996	0.59269	95	0.15786	0.206645
46	0.55513	0.58874	96	0.149375	0.198689
47	0.553202	0.58675	97	0.148873	0.198196
48	0.545711	0.578828	98	0.140123	0.190289
49	0.539075	0.57143	99	0.139446	0.18959
50	0.537544	0.570022	100	0.136725	0.186873



الشكل 2 : يوضح دالة البقاء للمرضى لبيانات المراقبة من اليمين

من خلال الجدول (5) والشكل (2) نلاحظ ان دالة البقاء معاكسة لدالة الخطر الموضح في الجدول (6) والشكل (3) اي كلما كان الخطر عالي كلما كان بقاء المرضى على قيد الحياة اقل .

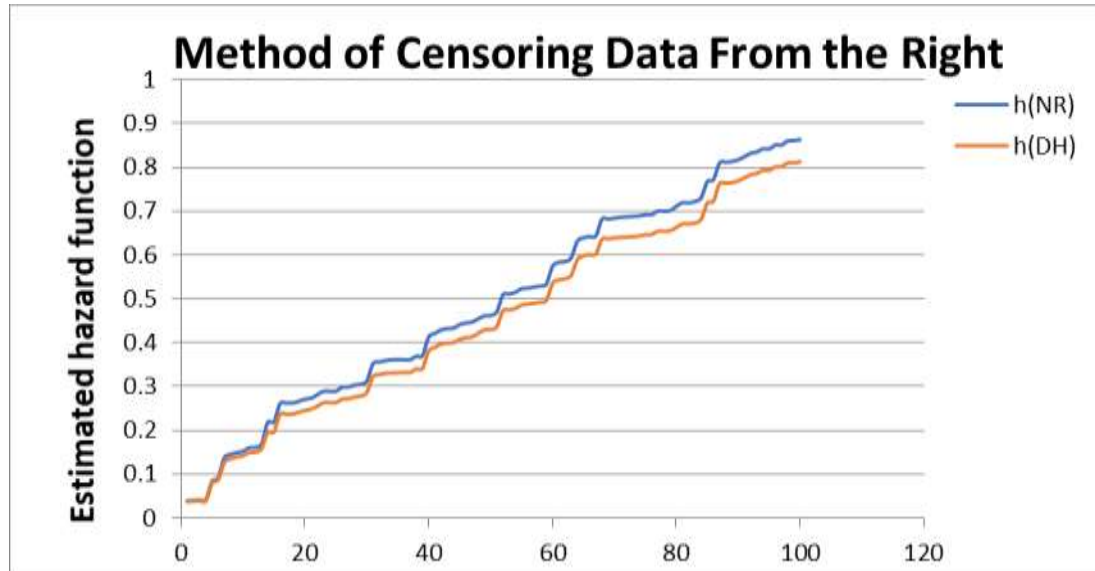
وان دالة الخطر لنموذج كوكس للمقدرات المستعملة لبيانات المراقبة من اليمين موضحة في الجدول والمخطط الاتي .

الجدول 6 : يوضح دالة الخطر لأنموذج Cox لبيانات المراقبة من اليمين

t	h(NR)	h(DH)	t	h(NR)	h(DH)
1	0.039	0.037	51	0.470	0.436
2	0.040	0.038	52	0.509	0.473
3	0.041	0.039	53	0.511	0.475
4	0.041	0.039	54	0.515	0.478
5	0.083	0.079	55	0.523	0.486
6	0.092	0.087	56	0.524	0.488
7	0.137	0.127	57	0.527	0.490
8	0.145	0.135	58	0.529	0.492
9	0.149	0.139	59	0.533	0.496
10	0.152	0.142	60	0.574	0.535
11	0.160	0.149	61	0.583	0.543
12	0.161	0.150	62	0.585	0.546
13	0.168	0.158	63	0.593	0.553
14	0.218	0.195	64	0.631	0.589
15	0.219	0.196	65	0.639	0.598
16	0.261	0.237	66	0.642	0.601

17	0.262	0.237	67	0.643	0.602
18	0.262	0.237	68	0.682	0.636
19	0.266	0.241	69	0.682	0.637
20	0.271	0.245	70	0.684	0.639
21	0.273	0.248	71	0.686	0.640
22	0.281	0.255	72	0.687	0.641
23	0.289	0.263	73	0.688	0.642
24	0.289	0.263	74	0.689	0.643
25	0.289	0.263	75	0.692	0.646
26	0.298	0.271	76	0.692	0.647
27	0.299	0.272	77	0.700	0.654
28	0.303	0.276	78	0.700	0.654
29	0.305	0.278	79	0.701	0.655
30	0.312	0.286	80	0.711	0.663
31	0.352	0.323	81	0.719	0.671
32	0.356	0.327	82	0.719	0.671
33	0.359	0.330	83	0.722	0.673
34	0.361	0.331	84	0.730	0.681
35	0.361	0.331	85	0.767	0.718
36	0.361	0.332	86	0.772	0.723
37	0.361	0.332	87	0.810	0.763
38	0.369	0.340	88	0.812	0.764
39	0.370	0.341	89	0.813	0.765
40	0.413	0.381	90	0.817	0.769
41	0.421	0.389	91	0.824	0.776
42	0.429	0.397	92	0.832	0.783
43	0.432	0.399	93	0.835	0.786
44	0.433	0.400	94	0.842	0.793
45	0.441	0.407	95	0.842	0.793
46	0.445	0.411	96	0.851	0.801
47	0.447	0.413	97	0.851	0.802

48	0.454	0.421	98	0.860	0.810
49	0.461	0.429	99	0.861	0.810
50	0.462	0.430	100	0.863	0.813



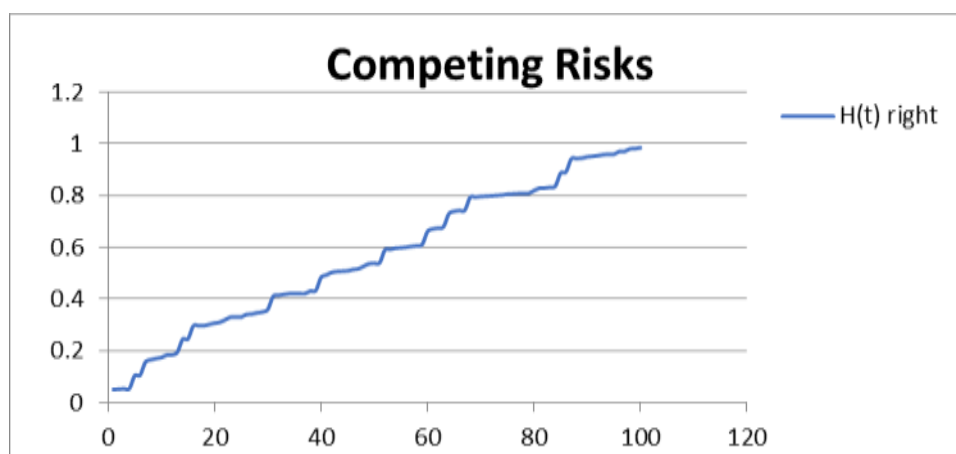
الشكل 3 : يوضح دالة الخطر لأنموذج Cox لبيانات المراقبة من اليمين

بعد ذلك تم تقدير المخاطرة التنافسية من خلال دالة الخطر التراكمي (CIF) لبيانات المراقبة من اليمين والذي يعرف على أنه احتمال وقوع الحدث في أي نقطة زمنية بين خط الأساس والوقت t ، حيث تقدر (CIF) في الوقت t من خلال نسبة الأشخاص الذين حصل لهم الحدث (الموت) مقسوم على عدد الإجمالي للأفراد المعرضين للخطر خلال t_i ، ومع زيادة الوقت يزيد CIF من الصفر إلى النسبة الإجمالية للأحداث والموضحة في الجدول الآتي :

جدول (7) : يوضح مقدرات المخاطرة التنافسية لبيانات المراقبة من اليمين

t	$\hat{H}(t)$	t	$\hat{H}(t)$	t	$\hat{H}(t)$	t	$\hat{H}(t)$
1	0.05	26	0.340326	51	0.539708	76	0.806176
2	0.051414	27	0.341755	52	0.589708	77	0.807477
3	0.052829	28	0.346705	53	0.592183	78	0.807529
4	0.053324	29	0.350006	54	0.597134	79	0.808944
5	0.103324	30	0.359907	55	0.598235	80	0.818944
6	0.105225	31	0.409907	56	0.600235	81	0.828845
7	0.155225	32	0.414857	57	0.603568	82	0.829175

8	0.165225	33	0.418158	58	0.606043	83	0.83165
9	0.170175	34	0.420633	59	0.610994	84	0.836551
10	0.173509	35	0.420693	60	0.660994	85	0.886551
11	0.18341	36	0.420802	61	0.670994	86	0.891551
12	0.184235	37	0.420959	62	0.674294	87	0.941551
13	0.194136	38	0.43086	63	0.679195	88	0.943201
14	0.244136	39	0.43286	64	0.729195	89	0.944851
15	0.24555	40	0.48286	65	0.739195	90	0.949802
16	0.29555	41	0.49286	66	0.742146	91	0.951703
17	0.29665	42	0.50286	67	0.743136	92	0.954604
18	0.296788	43	0.50616	68	0.793136	93	0.957904
19	0.301788	44	0.507398	69	0.794236	94	0.959805
20	0.306788	45	0.509099	70	0.796711	95	0.960007
21	0.310088	46	0.51405	71	0.798125	96	0.970007
22	0.319989	47	0.516525	72	0.799226	97	0.970667
23	0.32989	48	0.526426	73	0.800876	98	0.980667
24	0.330158	49	0.536327	74	0.801976	99	0.981576
25	0.330425	50	0.538307	75	0.805276	100	0.98491



شكل (4) : يوضح دالة الخطر التراكمي للمخاطرة التنافسية لبيانات المراقبة من اليمين

من خلال الجدول (7) الخاص بالمخاطرة التنافسية وبالمقارنة مع الجدول (6) الخاص بدالة الخطر لأنموذج Cox والاشكال التابعة لها نلاحظ زيادة المخاطرة التنافسية عن دالة الخطر للمقدرات نيوتن رافسون (NR) وداون هيل (DH) والذي يشير الى تأثير المتغيرات المشتركة في زيادة المخاطرة على دالة الخطر على دالة الخطر .

4. الاستنتاجات

1. اظهرت خوارزمية داون هيل (DH) تفوقها في تقدير دالة الخطر لنموذج كوكس لكلا الحالتين من المعلمات المفترضة ولنماذج مختلفة من الاوساط الحسابية والانحرافات المعيارية ولأحجام مختلفة من العينات في المحاكاة .
2. ظهور تطابق في نتائج الجانب التطبيقي حيث اظهر الجانب التطبيقي تفوق خوارزمية داون هيل (DH) في التقدير عن خوارزمية نيوتن رافسون (NR) .
3. من خلال تجارب المحاكاة تم ملاحظة تناقص متوسط مربع الخطأ (MRSE) لدالة الخطر لنموذج كوكس كلما زاد حجم العينة .
4. من خلال تقدير المخاطر التنافسية للبيانات الحقيقية تم توضيح تأثير المتغيرات المشتركة في زيادة الخطر على دالة الخطر لنموذج كوكس .
5. في الجانب التطبيقي نلاحظ زيادة دالة الخطر لنموذج كوكس مع مرور الوقت مما يؤدي الى انخفاض دالة البقاء .

المصادر

- 1- Andrade, C. (2023). "Survival analysis, Kaplan-Meier curves, and Cox regression: basic concepts". Indian Journal of Psychological Medicine, 02537176231176986. Vol. 45, No. 4, pp.434–435 .
- 2- Teoh, S. H. (2008). "Using Kaplan Meier and cox regression in survival analysis: An example". Teoh Sian . Hoon. ESTEEM Academic Journal, Vol. 4, No. 2, pp.3–14 .
- 3- Smith, T., and Smith, B. (2003). "Kaplan Meier and Cox proportional hazards modeling: hands on survival analysis". SAS® Users Group International Proc. Seattle, Washington. Vol. 6, No. 2, pp.1–14 .
- 4- Tang, Z., Zhou, C., Jiang, W., Zhou, W., Jing, X., Yu, J., and Sheng, B. (2013). "Analysis of significant factors on cable failure using the cox proportional hazard model". IEEE Transactions on Power Delivery, Vol. 29, No. 2, pp.951–957 .
- 5- Scheike, T. H., and Sun, Y. (2007). "Maximum likelihood estimation for tied survival data under Cox regression model via EM-algorithm". Lifetime data analysis, Vol. 13, No. 4, pp.399–420 .
- 6- Thackham, M., and Ma, J. (2020). "On maximum likelihood estimation of the semi-parametric Cox model with time-varying covariates". Journal of Applied Statistics, Vol. 47, No. 9, pp.1511–1528 .
- 7- Yalçinkaya, A., Şenoğlu, B., and Yolcu, U. (2018). "Maximum likelihood estimation for the parameters of skew normal distribution using genetic algorithm". Swarm and Evolutionary Computation, Vol. 38, No. 16, pp.127–138 .

- 8- Akram, S., and Ann, Q. U. (2015). "Newton raphson method". International Journal of Scientific & Engineering Research, Vol. 6, No. 7, pp.1748–1752 .
- 9- Casella, F., and Bachmann, B. (2021). "On the choice of initial guesses for the Newton-Raphson algorithm". Applied Mathematics and Computation, Vol. 3, No. 18, pp.1–38 .
- 10- Fajfar, I., Puhán, J., and Bűrmen, Á. (2017). "Evolving a Nelder–Mead algorithm for optimization with genetic programming". Evolutionary computation, Vol. 25, No. 3, pp.351–373 .
- 11- Galántai, A. (2021). "A convergence analysis of the Nelder-Mead simplex method". Acta Polytech. Hung, Vol. 18, No. 5, pp.93–105
- 12- Mehta, V. K. (2020). "Improved nelder–mead algorithm in high dimensions with adaptive parameters based on chebyshev spacing points". Engineering Optimization, Vol. 52, No. 10, pp.1814–1828 .
- 13- Fan, J., Lin, H., and Zhou, Y. (2006). "Local partial-likelihood estimation for lifetime data". Vol. 34, No. 1, pp.290–325
- 14- Sinha, D., Ibrahim, J. G., and Chen, M. H. (2003). "A Bayesian justification of Cox's partial likelihood". Biometrika, Vol.90, No 3, pp.629-641.
- 15- Cox, D. R. (1975). "Partial likelihood. Biometrika", Vol.62, No 2, pp.269-276.
- 16- Groeneboom, P., Maathuis, M. H., and Wellner, J. A. (2008). "Current status data with competing risks: Limiting distribution of the MLE". Annals of statistics, Vol.36, No 3, pp.1064-1089.
- 17- Kalbfleisch, J. D., and Prentice, R. L. (1973). "Marginal likelihoods based on Cox's regression and life model". Biometrika, Vol.60, No 2, pp.267-278.
- 18- Sildnes, B. (2015). "Modelling of Dependent Competing Risks and Semi-Competing Risks by Means of First Passage Times of Gamma Processes". (Master's thesis, NTNU), Vol.85, No 6, pp.1-258.
- 19- Hudgens, M. G., Li, C., and Fine, J. P. (2014). "Parametric likelihood inference for interval censored competing risks data". Biometrics, Vol.70, No 1, pp.1-9.
- 20- Mao, L., and Lin, D. Y. (2017). "Efficient estimation of semiparametric transformation models for the cumulative incidence of competing risks". Journal of the Royal Statistical Society Series B: Statistical Methodology, Vol.79, No 2 , pp.573-587.
- 21- Thackham, M., and Ma, J. (2020). "On maximum likelihood estimation of the semi-parametric Cox model with time-varying covariates". Journal of Applied Statistics, Vol.47, No 9 , pp.1511-1528.
- 22- Zeng, D., Mao, L., and Lin, D. Y. (2016). "Maximum likelihood estimation for semiparametric transformation models with interval-censored data". Biometrika, Vol.103, No 2, pp. 253-271.
- 23- Guerrero, A., Del Campo, L., Piscaglia, F., Scheiner, B., Han, G., Violi, F., and Girleanu, I. (2023). "Anticoagulation improves survival in patients with cirrhosis and portal vein thrombosis": The IMPORTAL competing-risk meta-analysis. Journal of Hepatology , Vol.79, No 1,

STUDY OF CHEMICAL AND HEALTH PROBLEMS BY KUFFI ISSA TRANSFORM

Ahmad Issa¹, Emad A. Kuffi²

¹Department of Mathematics, Karabuk University, Karabuk,
Turkey.

²Department of Mathematics, Mustansiriyah University,
Baghdad, Iraq.

Mathematics Subject Classification (2010): 34A30, 44A15

STUDY OF CHEMICAL AND HEALTH PROBLEMS BY KUFFI ISSA TRANSFORM

Ahmad Issa¹, Emad A. Kuffi²

¹Department of Mathematics, Karabuk University, Karabuk, Turkey.

²Department of Mathematics, Mustansiriyah University, Baghdad, Iraq.

Mathematics Subject Classification (2010): 34A30, 44A15

Keywords and phrases: Kuffi Issa Transform, Inverse Kuffi Issa Transform, Models in Chemical Reactions, Models in Health Sciences, Ordinary Differential Equation.

Abstract

This paper introduces the Kuffi-Issa Transform (KIT), a new mathematical technique developed to deal with problems in chemistry, health, physics, and other sciences. To test the validity of this transform, three chemical and health examples are presented. The results show that this transform is a powerful tool for solving differential equations encountered in chemical and health problems.

1 Introduction

Integral transformations are used to solve differential equations, and are widely used in physics, engineering, and applied mathematics [1, 2, 3]. Through these transformations, a specific function in a domain such as time is transformed into another domain such as frequency. To solve these equations, several transformations have been introduced, including Laplace transform [4], Fourier transform [5], Elzaki transform [6], and Aboodh Transform [7]. Recently, new transformations have been introduced, including complex SEE transform [8], Emad-Sara transform [9], Mayan transform [10], and Jafari transform [11]. In this paper, we will introduce a new transformation (Kuffi Issa transform) to solve differential equations.

2 Preliminaries

2.1 Definitions and Theorems of Kuffi Issa Transform

Definition 2.1 The Kuffi Issa transform of $f(t)$ denoted by $KI[.]$ is given by

$$KI[f(t)] = (uv)^{i\beta} \int_{t=0}^{\infty} f(v^{i\gamma} t) e^{-u^{i\alpha} t} dt = F(u, v), \quad (1)$$

where u, v are parameters; $t \geq 0; \alpha, \beta, \gamma \in \mathbb{Z}$.

Definition 2.2 The Inverse Kuffi Issa transform of $F(u, v)$ denoted by $(KI)^{-1}[.]$ is given by

$$(KI)^{-1}[KI[f(t)]] = \frac{1}{2\pi i} \lim_{\epsilon \rightarrow 0} \int_{\mu - i\epsilon}^{\mu + i\epsilon} (uv)^{i\beta} v^{i\gamma} e^{\frac{u^{i\alpha}}{v^{i\alpha}} F(u, v)} du = f(t). \quad (2)$$

Theorem 2.3 (Linearity of KI Transform)

If $KI[f_1(t)] = F_1(u, v)$ and $KI[f_2(t)] = F_2(u, v)$, then

$$KI[c_1 f_1(t) \mp c_2 f_2(t)] = c_1 F_1(u, v) \mp c_2 F_2(u, v).$$

Proof. By using Eq (1) we obtain

$$\begin{aligned} KI[c_1 f_1(t) \mp c_2 f_2(t)] &= (uv)^{i\beta} \int_0^{\infty} [c_1 f_1(v^{i\gamma} t) \mp c_2 f_2(v^{i\gamma} t)] e^{-u^{i\alpha} t} dt \\ &= c_1 (uv)^{i\beta} \int_0^{\infty} f_1(v^{i\gamma} t) e^{-u^{i\alpha} t} dt \mp c_2 (uv)^{i\beta} \int_0^{\infty} f_2(v^{i\gamma} t) e^{-u^{i\alpha} t} dt \\ &= c_1 F_1(u, v) \mp c_2 F_2(u, v). \end{aligned}$$

Theorem 2.4 Let C be a scalar and $f(t) = C$, then $KI[f(t)] = C u^{i(\beta-\alpha)} v^{i\beta}$

Proof. By using Eq (1) we obtain

$$KI[C] = (uv)^{i\beta} \int_0^\infty C e^{-u^{i\alpha} t} dt = C(uv)^{i\beta} \frac{e^{-u^{i\alpha} t}}{-u^{i\alpha}} \Big|_0^\infty = C u^{i(\beta-\alpha)} v^{i\beta}$$

Theorem 2.5 If $f(t) = t$, then $KI[f(t)] = u^{i(\beta-2\alpha)} v^{i(\beta+\gamma)}$

Proof. By using Eq (1) we obtain

$$\begin{aligned} KI[t] &= (uv)^{i\beta} \int_0^\infty v^{i\gamma} t e^{-u^{i\alpha} t} dt = u^{i\beta} v^{i(\beta+\gamma)} \int_0^\infty t e^{-u^{i\alpha} t} dt \\ &= u^{i(\beta-\alpha)} v^{i(\beta+\gamma)} \frac{e^{-u^{i\alpha} t}}{-u^{i\alpha}} \Big|_0^\infty = u^{i(\beta-2\alpha)} v^{i(\beta+\gamma)} \end{aligned}$$

Theorem 2.6 If $f(t) = e^{at}$, then $KI[f(t)] = \frac{(uv)^{i\beta}}{u^{i\alpha} - av^{i\gamma}}$

Proof. By using Eq (1) we obtain

$$\begin{aligned} KI[e^{at}] &= (uv)^{i\beta} \int_0^\infty e^{a(v^{i\gamma} t)} e^{-u^{i\alpha} t} dt = (uv)^{i\beta} \int_0^\infty e^{-(u^{i\alpha} - av^{i\gamma})t} dt \\ &= \frac{(uv)^{i\beta}}{-(u^{i\alpha} - av^{i\gamma})} e^{-(u^{i\alpha} - av^{i\gamma})t} \Big|_0^\infty = \frac{(uv)^{i\beta}}{u^{i\alpha} - av^{i\gamma}} \end{aligned}$$

Theorem 2.7 If $f(t) = e^{-at}$, then $KI[f(t)] = \frac{(uv)^{i\beta}}{u^{i\alpha} + av^{i\gamma}}$

Proof. By using Eq (1) we obtain

$$\begin{aligned} KI[e^{-at}] &= (uv)^{i\beta} \int_0^\infty e^{-a(v^{i\gamma} t)} e^{-u^{i\alpha} t} dt = (uv)^{i\beta} \int_0^\infty e^{-(u^{i\alpha} + av^{i\gamma})t} dt \\ &= \frac{(uv)^{i\beta}}{-(u^{i\alpha} + av^{i\gamma})} e^{-(u^{i\alpha} + av^{i\gamma})t} \Big|_0^\infty = \frac{(uv)^{i\beta}}{u^{i\alpha} + av^{i\gamma}} \end{aligned}$$

Theorem 2.8 If $f(t) = e^{ait}$, then $KI[f(t)] = \frac{(uv)^{i\beta}}{u^{i\alpha} - aiv^{i\gamma}}$

Proof. By using Eq (1) we obtain

$$\begin{aligned} KI[e^{ait}] &= (uv)^{i\beta} \int_0^\infty e^{ai(v^{i\gamma} t)} e^{-u^{i\alpha} t} dt = (uv)^{i\beta} \int_0^\infty e^{-(u^{i\alpha} - aiv^{i\gamma})t} dt \\ &= \frac{(uv)^{i\beta}}{-(u^{i\alpha} - aiv^{i\gamma})} e^{-(u^{i\alpha} - aiv^{i\gamma})t} \Big|_0^\infty = \frac{(uv)^{i\beta}}{u^{i\alpha} - aiv^{i\gamma}} \end{aligned}$$

Theorem 2.9 If $f(t) = e^{-ait}$, then $KI[f(t)] = \frac{(uv)^{i\beta}}{u^{i\alpha} + ai v^{i\gamma}}$

Proof. By using Eq (1) we obtain

$$\begin{aligned} KI[e^{-ait}] &= (uv)^{i\beta} \int_0^\infty e^{-ai(v^{i\gamma}t)} e^{-u^{i\alpha}t} dt = (uv)^{i\beta} \int_0^\infty e^{-(u^{i\alpha} + ai v^{i\gamma})t} dt \\ &= \frac{(uv)^{i\beta}}{-(u^{i\alpha} + ai v^{i\gamma})} e^{-(u^{i\alpha} + ai v^{i\gamma})t} \Big|_0^\infty = \frac{(uv)^{i\beta}}{u^{i\alpha} + ai v^{i\gamma}} \end{aligned}$$

Theorem 2.10 If $f(t) = \cos at$, then $KI[f(t)] = \frac{u^{(\alpha+\beta)i} v^{i\beta}}{u^{2i\alpha} + a^2 v^{2i\gamma}}$

Proof. By using Eq (1) we obtain

$$\begin{aligned} KI[\cos at] &= KI\left[\frac{e^{ait} + e^{-ait}}{2}\right] = \frac{1}{2}[KI[e^{ait}] + KI[e^{-ait}]] = \frac{1}{2}\left[\frac{(uv)^{i\beta}}{u^{i\alpha} - ai v^{i\gamma}} + \frac{(uv)^{i\beta}}{u^{i\alpha} + ai v^{i\gamma}}\right] \\ &= \frac{u^{(\alpha+\beta)i} v^{i\beta}}{u^{2i\alpha} + a^2 v^{2i\gamma}} \end{aligned}$$

Theorem 2.11 If $f(t) = \sin at$, then $KI[f(t)] = \frac{au^{i\beta} v^{i(\beta+\gamma)}}{u^{2i\alpha} + a^2 v^{2i\gamma}}$

Proof. By using Eq (1) we obtain

$$\begin{aligned} KI[\sin at] &= KI\left[\frac{e^{ait} - e^{-ait}}{2i}\right] = \frac{1}{2i}[KI[e^{ait}] - KI[e^{-ait}]] = \frac{1}{2i}\left[\frac{(uv)^{i\beta}}{u^{i\alpha} - ai v^{i\gamma}} - \frac{(uv)^{i\beta}}{u^{i\alpha} + ai v^{i\gamma}}\right] \\ &= \frac{au^{i\beta} v^{i(\beta+\gamma)}}{u^{2i\alpha} + a^2 v^{2i\gamma}} \end{aligned}$$

Theorem 2.12 If $f(t) = \cosh at$, then $KI[f(t)] = \frac{u^{(\alpha+\beta)i} v^{i\beta}}{u^{2i\alpha} - a^2 v^{2i\gamma}}$

Proof. By using Eq (1) we obtain

$$\begin{aligned} KI[\cosh at] &= KI\left[\frac{e^{at} + e^{-at}}{2}\right] = \frac{1}{2}[KI[e^{at}] + KI[e^{-at}]] = \frac{1}{2}\left[\frac{(uv)^{i\beta}}{u^{i\alpha} - av^{i\gamma}} + \frac{(uv)^{i\beta}}{u^{i\alpha} + av^{i\gamma}}\right] \\ &= \frac{u^{(\alpha+\beta)i} v^{i\beta}}{u^{2i\alpha} - a^2 v^{2i\gamma}} \end{aligned}$$

Theorem 2.13 If $f(t) = \sinh at$, then $KI[f(t)] = \frac{au^{i\beta} v^{i(\beta+\gamma)}}{u^{2i\alpha} - a^2 v^{2i\gamma}}$

Proof. By using Eq (1) we obtain

$$\begin{aligned}
 KI[\sinh at] &= KI \left[\frac{e^{at} - e^{-at}}{2} \right] = \frac{1}{2} [KI[e^{at}] - KI[e^{-at}]] = \frac{1}{2} \left[\frac{(uv)^{i\beta}}{u^{i\alpha} - av^{i\gamma}} - \frac{(uv)^{i\beta}}{u^{i\alpha} + av^{i\gamma}} \right] \\
 &= \frac{au^{i\beta} v^{i(\beta+\gamma)}}{u^{2i\alpha} - a^2 v^{2i\gamma}}
 \end{aligned}$$

2.2 Kuffi Issa Transform of derivatives of $f(t)$

$$\begin{aligned}
 KI[f'(t)] &= (uv)^{i\beta} \int_0^\infty f'(v^{i\gamma} t) e^{-u^{i\alpha} t} dt, \\
 &= (uv)^{i\beta} \left[e^{-u^{i\alpha} t} \frac{f(v^{i\gamma} t)}{v^{i\gamma}} \Big|_0^\infty - \int_0^\infty f(v^{i\gamma} t) e^{-u^{i\alpha} t} \left(\frac{-u^{i\alpha}}{v^{i\gamma}} \right) dt \right] \\
 &= \frac{u^{i\alpha}}{v^{i\gamma}} F - f(0) u^{i\beta} v^{i(\beta-\gamma)}
 \end{aligned}$$

3 Applications of Kuffi Issa Transform

Now we will provide a solution to some chemical and health problems using Kuffi Issa Transform

Application 3.1 Consider the following Differential Equation

$$\frac{dc}{dt} = -k_0 c, \quad c(0) = c_0 \quad (3)$$

By taking KI transform of Eq (3), we obtain

$$KI \left[\frac{dc}{dt} \right] = KI[-k_0 c]$$

$$\frac{u^{i\alpha}}{v^{i\gamma}} F - c_0 u^{i\beta} v^{i(\beta-\gamma)} = -k_0 u^{i(\beta-\alpha)} v^{i\beta},$$

$$F = (-k_0 u^{i(\beta-\alpha)} v^{i\beta} + c_0 u^{i\beta} v^{i(\beta-\gamma)}) \cdot \frac{v^{i\gamma}}{u^{i\alpha}},$$

$$F = -k_0 u^{i(\beta-2\alpha)} v^{i(\beta+\gamma)} + c_0 u^{i(\beta-\alpha)} v^{i\beta},$$

The analytical solution of (3) is

$$c(t) = F^{-1} = -k_0 t + c_0$$

Application 3.2 Consider the following Differential Equation

$$\frac{dc}{dt} = -k_1 c, c(0) = c_0, k_1 > 0 \quad (4)$$

By taking KI transform of Eq (4), we obtain

$$KI\left[\frac{dc}{dt}\right] = KI[-k_1 c]$$

$$\frac{u^{i\alpha}}{v^{i\gamma}} F - c_0 u^{i\beta} v^{i(\beta-\gamma)} = -k_1 F,$$

$$F = (c_0 u^{i\beta} v^{i(\beta-\gamma)}) \cdot \frac{v^{i\gamma}}{u^{i\alpha} + k_1 v^{i\gamma}},$$

$$F = c_0 \frac{(uv)^{i\beta}}{u^{i\alpha} + k_1 v^{i\gamma}},$$

The analytical solution of (4) is

$$c(t) = F^{-1} = c_0 e^{-k_1 t}$$

Application 3.3 Consider the following Differential Equation

$$\frac{dy}{dt} + \frac{y(t)}{200} = \frac{3}{4}, y(0) = 20, \quad (5)$$

By taking KI transform of Eq (5), we obtain

$$KI\left[\frac{dy}{dt} + \frac{y(t)}{200}\right] = KI\left[\frac{3}{4}\right]$$

$$\frac{u^{i\alpha}}{v^{i\gamma}} F - y(0)u^{i\beta}v^{i(\beta-\gamma)} + \frac{1}{200}F = \frac{3}{4}u^{i(\beta-\alpha)}v^{i\beta},$$

$$F = \frac{4000(uv)^{i\beta} + 150u^{i(\beta-\alpha)}v^{i(\beta+\gamma)}}{200u^{i\alpha} + v^{i\gamma}},$$

$$F = 150u^{i(\beta-\alpha)}v^{i\beta} - \frac{130(uv)^{i\beta}}{u^{i\alpha} + \frac{1}{200}v^{i\gamma}},$$

The analytical solution of (5) is

$$y(t) = F^{-1} = 150 - 130e^{\frac{-t}{200}}$$

4 Conclusion

In this paper, a new transformation (Kuffi-Issa Transform) is constructed and applied to chemical and health problems. The main advantage of KIT transformation is that it gives solutions with less effort without dealing with long and complicated calculations. These complex powers in this transformation are what gave ease in simplifying the solution, as well as generality to many integral transformations with two parameters. In addition, this transform has been proven to solve differential equations. In future studies, we can use this transform to deal with partial differential equations, and difference equations.

References

- [1] Issa, A., Kuffi, E. and Düz, M., A Further Generalization of the General Polynomial Transform and its Basic Characteristics and Applications, Journal of University of Anbar for Pure Science. 17(2) (2023) 338-342.
- [2] Kumar, P.S. and Viswanathan, A., Application of Mahgoub transform to mechanics, electrical circuit problems, International Journal of Science and Research. 7(7) (2018) 195-197.
- [3] Ahmadi, S.A.P., Hosseinzadeh, H. and Cherati, A.Y., A new integral transform for solving higher order linear ordinary Laguerre and Hermite differential equations, International Journal of Applied and Computational Mathematics, 5 (2019) 1-7.
- [4] Pavani, P.V., Priya, U.L. and Reddy, B.A., Solving differential equations by using Laplace transforms, Int. J. Res. Anal. Rev. 5 (2018) 1796-1799.
- [5] Düz, M., Avezov, S. and Issa, A., Solutions to Differential-Differential Difference Equations with Variable Coefficients by Using Fourier Transform Method, Süleyman Demirel University Faculty of Arts and Science Journal of Science. 18(3) (2023) 259-267.
- [6] Elzaki, T.M., The new integral transform Elzaki transform, Global Journal of pure and applied mathematics. 7(1) (2011) 57-64.
- [7] Aboodh, K.S., The New Integral Transform' Aboodh Transform, Global journal of pure and Applied mathematics. 9(1) (2013) 35-43.
- [8] Mansour, E.A., Mehdi, S. and Kuffi, E.A., The new integral transform and its applications. International Journal of Nonlinear Analysis and Applications. 12(2) (2021) 849-856.
- [9] Maktoof, S.F., Kuffi, E. and Abbas, E.S., "Emad-Sara Transform" a new integral transform, Journal of Interdisciplinary Mathematics. 24(7) (2021) 1985-1994.

- [10] Mansour, E.A. and Kuffi, E.A., The Mayan Transform: A Novel Integral Transform of Complex Power Parameters and Applications to Neutrosophic Functions, Full Length Article. 23(1) (2023) 323-334.
- [11] El-Mesady, A.I., Hamed, Y.S. and Alsharif, A.M., Jafari transformation for solving a system of ordinary differential equations with medical application, Fractal and Fractional. 5(3) (2021) 130.
- [12] Emad Kuffi, Eman Mansour, Ahmad Issa., SEE Transform Technique in Control Theory, Journal of Kufa for Mathematics and Computer. 11(2) (2024) 147-149.

العلاقة بين علوم الرياضيات ، الإحصاء والحاسبات

Mathematics, Statistics & Computer Science (MSCS)

الأستاذ الدكتور فاضل عبدالعباس العابدي

الجامعة الإسلامية / النجف الأشرف

abidy.fadhil@iunajaf.edu.iq

العلاقة بين علوم الرياضيات ، الإحصاء والحاسبات
Mathematics, Statistics & Computer Science (MSCS)



الأستاذ الدكتور فاضل عبدالعباس العابدي
الجامعة الإسلامية / النجف الأشرف
abidy.fadhil@iunajaf.edu.iq

المخلص:

بعد إجراء متابعة لأستعراض البحوث والمقالات التي تبحث عن العلاقة بين العلوم الثلاث (الرياضيات ، الإحصاء والحاسوب) لاحظنا إن علم الرياضيات يركز على النظريات الرياضية بينما يركز علم الإحصاء على استخدام تلك النظريات لجمع البيانات وتحليلها وأظهار النتائج، الإحصاء ظهر كأحد فروع الرياضيات يهتم بوضع اسس لجمع البيانات بشكل علمي موثوق وكذلك يستخدم القوانين الرياضية ومنحنيات التوزيع الاحصائية لتفسير هذه البيانات ، المهارات الرياضية ضرورية للإحصائيين لأنها تمنح القدرة على إجراء حسابات معقدة وأيضاً لتحديد الطرائق المناسبة لمشروع معين بالإضافة إلى فهم الأساليب الإحصائية نحتاج إلى معرفة حساب التفاضل والتكامل والجبر الخطي والاحتمالات.

كذلك الرياضيات تلعب دورًا حاسمًا في علوم الحاسوب، حيث تشكل الركيزة الأساسية التي يُبنى عليها هذا العلم حيث تمكّن المعرفة الجيدة بالرياضيات المبرمجين وعلماء الحاسوب من فهم كيفية عمل أجهزة الكمبيوتر ووحدات التحكم المنطقية، وتطوير خوارزميات أفضل والعمل على التشفير وحماية البيانات وأن الأساس القوي في الرياضيات المتقدمة ضروري لدراسة دقيقة لعلوم الكمبيوتر.

وقد تم استنتاج وجود علاقة مهمة وثيقة بين علوم الإحصاء وعلوم الحاسوب نتيجة دخول الأساليب الإحصائية في السنوات الأخيرة علم البيانات من خلال (Data Science/Gig Data /Data analysis) حيث أصبحت جانبًا مهمًا من الحوسبة العلمية في مجالات عديدة منها تعلم الآلة "الذكاء الاصطناعي" التي تحتاج إلى إنشاء تطبيقات برمجية متخصصة وبلغات حاسوب متقدمة مثل Python ,R-packages.

Abstract:

after reviewing research studies and articles exploring the relationship between the three sciences (Mathematics, Statistics, and Computer Science), we observed that Mathematics focuses on mathematical theories, while Statistics emphasizes the application of these theories to collect, analyze, and present data. Statistics emerged as a branch of Mathematics concerned with establishing reliable scientific methods for data collection and utilizing mathematical formulas and statistical distribution curves to interpret this data. Mathematical skills are essential for statisticians as they provide the ability to perform complex calculations, determine appropriate methods for specific projects, and understand statistical techniques, which require knowledge of calculus, linear algebra, and probability.

Similarly, Mathematics plays a crucial role in Computer Science, forming the foundation upon which this field is built. A strong mathematical background enables programmers and computer scientists to understand how computers and logical control units operate, develop better algorithms, work on encryption, and ensure data security. A solid foundation in advanced mathematics is essential for an in-depth study of Computer Science.

It has also been concluded that there is a significant and strong relationship between Statistics and Computer Science due to the increasing use of statistical methods in data science over the past few years (such as Data Science, Big Data, and Data Analysis). This has become an essential aspect of scientific computing in various fields, including Machine Learning and Artificial Intelligence, which require the development of specialized software applications using advanced programming languages such as R-packages and Python.

Keywords: Data analysis, Computer science, Mathematics, Statistics.

1. المقدمة

تلعب الرياضيات والإحصاء وعلوم الحاسوب دورًا حيويًا في التطور العلمي والتكنولوجي، حيث تُشكل هذه العلوم الثلاثة أساسًا متينًا للعديد من المجالات الحديثة، مثل الذكاء الاصطناعي، وتحليل البيانات، والتشفير، وتطوير

الخوارزميات. يركز علم الرياضيات على النظريات والمفاهيم المجردة، بينما يهتم علم الإحصاء بجمع البيانات وتحليلها من خلال الأساليب الرياضية، أما علوم الحاسوب فتستخدم هذه الأسس لبناء أنظمة متطورة تحاكي الذكاء البشري وتساهم في تحسين الأداء التقني لمختلف التطبيقات. في هذا البحث، سنستعرض العلاقة الوثيقة بين هذه العلوم الثلاثة، وأهمية تكاملها في تحقيق تقدم علمي وابتكارات تقنية متطورة مع علاقتها بعلوم تقنية المعلومات (IT) وتحليل البيانات .

تُعد الرياضيات، الإحصاء، وعلوم الحاسوب من العلوم الأساسية التي تتكامل فيما بينها لتشكّل حجر الأساس في العديد من التطبيقات العلمية والتكنولوجية الحديثة. فالرياضيات توفر المفاهيم والنظريات الأساسية التي يعتمد عليها كل من الإحصاء والحاسبات، حيث يركز علم الرياضيات على دراسة البنى المجردة والأنماط العددية، بينما يعتمد علم الإحصاء على الأساليب الرياضية لتحليل البيانات واستخلاص النتائج، في حين تُسخر علوم الحاسوب هذه المفاهيم لبناء أنظمة برمجية متطورة تعتمد على تحليل البيانات، الذكاء الاصطناعي، والخوارزميات الفعالة. ومن خلال هذا التكامل، أصبحت هذه العلوم تلعب دورًا حاسمًا في التطورات التقنية الحديثة، مما يعزز أهمية دراستها وفهم العلاقة الوثيقة التي تربط بينها.

2. علاقة الرياضيات بعلم الإحصاء Ma & St [1,3,7,11]

إن علم الرياضيات يركز على النظريات الرياضية والأدوات المستخدمة في دراستها، بينما يركز علم الإحصاء على استخدام تلك الأدوات لجمع البيانات وتحليلها واستنتاج النتائج، الإحصاء ظهر كاحد فروع الرياضيات يهتم بوضع أسس لجمع البيانات بشكل علمي موثوق وكذلك يستخدم القوانين الرياضية ومنحنيات التوزيع الاحصائية لتفسير هذه البيانات كما يعتبر علم الإحصاء أداة من ادوات البحث العلمي يمكن استخدام قوانينها في جميع البحوث الطبية والنفسية وغيرها.

المهارات الرياضية ضرورية للإحصائيين لأنها تمنح القدرة على إجراء حسابات معقدة وأيضاً لتحديد الطرائق المناسبة لمشروع معين بالإضافة إلى فهم الأساليب الإحصائية نحتاج إلى معرفة حساب التفاضل والتكامل والجبر الخطي والاحتمالات.

1.2 التشابه والاختلاف بين الرياضيات والإحصاء

علم الإحصاء وعلم الرياضيات يتشابهان ببعض الجوانب، لكنهما يختلفان في المنهج والتطبيقات. وفيما يلي بعض الاختلافات الأساسية بينهما:

- a.** المنهج: يركز علم الرياضيات على دراسة الأنماط والعلاقات الرياضية، بينما يركز علم الإحصاء على تجميع البيانات وتحليلها وتفسيرها.
- b.** التطبيقات: يستخدم علم الرياضيات في مجالات مثل الفيزياء والهندسة والكمبيوتر وغيرها، بينما يستخدم علم الإحصاء في مجالات مثل الإدارية والاقتصادية، العلوم الاجتماعية، الزراعية والطب وغيرها.
- c.** الأدوات: يستخدم علم الرياضيات أدوات مثل الأعداد والمعادلات والمساحة والحساب، بينما يستخدم علم الإحصاء أدوات مثل الاحتمالات والتحليل الإحصائي والتصميم العيني.
- d.** الهدف: يهدف علم الرياضيات إلى دراسة الأنماط والقواعد الرياضية، بينما يهدف علم الإحصاء إلى تحليل البيانات وتوفير الإجابات الكمية على الأسئلة المتعلقة بالظواهر الاجتماعية والاقتصادية والطبية وغيرها.

2.2 مساهمة الرياضيات في علم الإحصاء

- a.** المعادلات التفاضلية : هناك أهمية واستخدامات كبيرة للمعادلات التفاضلية في علم الإحصاء من خلال نمذجة العمليات العشوائية لوصف تطور النظم التي تتأثر بالعشوائية، مثل الأسواق المالية أو الأنظمة البيولوجية. تُستخدم هذه المعادلات لنمذجة سلوك المتغيرات العشوائية عبر الزمن. بالإضافة إلى العمليات الديناميكية ، تحليل البقاء، تحليل الشبكات تُستخدم المعادلات التفاضلية في تحليل الشبكات، مثل الشبكات العصبية الاصطناعية، وأخيرا" لها دور في نمذجة الانتشار للأمراض والمواد الكيميائية.
- b.** التفاضل والتكامل هي أدوات رياضية أساسية لها دور حيوي في علم الإحصاء، حيث تساهم في تطوير الأساليب الإحصائية لتحليل البيانات (النماذج الخطية واللاخطية، التوزيعات العشوائية، نماذج السلاسل الزمنية ،) .
- c.** الجبر الخطي: الجبر الخطي هو الأساس الرياضي لتحليل الانحدار الخطي، حيث تُستخدم المصفوفات والمتجهات لحل أنظمة المعادلات الخطية و تحليل المكونات الرئيسية (PCA): تُستخدم تقنيات الجبر الخطي لاستخراج المكونات الرئيسية من البيانات وتقليل الأبعاد و تحليل القيم الذاتية وهو أمر مهم في تحليل التباين وتحليل العوامل.
- d.** المصفوفات طرائق التقدير: تُستخدم مصفوفات المعاملات في تقدير المعلمات في النماذج الإحصائية المعقدة.
- باختصار، التفاضل والتكامل والجبر الخطي هي أدوات رياضية حيوية تُمكن الإحصائيين من تطوير نماذج معقدة، تحليل البيانات بدقة، واستخلاص استنتاجات من البيانات التي قد تكون غير مرئية بدون هذه الأدوات الرياضية المتقدمة.

3. علاقة علم الإحصاء بعلوم الحاسوب St & CS [12]

في السنوات الأخيرة زاد الاهتمام بالبيانات من خلال الدخول الى علوم البيانات (Data Science) حيث أصبحت جانباً مهماً من الحوسبة العلمية في مجالات عديدة منها تعلم الآلة "الذكاء الاصطناعي" إلى إنشاء تطبيقات متخصصة وموجهة، هذا الاعتماد المتزايد على البيانات جعل الإحصاء موضوعاً مهماً للدراسة لجميع مناهج علوم الحاسوب ، حيث تسمح الأساليب الإحصائية لبرامج الحاسوب بتقديم تعميمات وتوقعات دقيقة بناءً على المعلومات المتاحة. وفهم الإحصاء يسهل فهم مفاهيم مثل (تعلم الآلة التعرف على الكلام استجابة المستخدم تحليل رسومات الحاسوب) حساب التفاضل والتكامل حساب التفاضل والتكامل هو فحص للتغيرات ومعدلات التغيير المستمرة ويتولى معرفة النتائج وخصائص التكاملات ومشتقات الدوال. ويستخدم في مجموعة من مجالات علم الحاسوب، بما في ذلك إنشاء الرسوم البيانية والمحاكاة والترميز والتشفير في التطبيقات وإنشاء الحلول الإحصائية وتصميم الخوارزميات وتحليلها.

ويمكن أن نلخص أن بين علوم الحاسوب والإحصاء علاقة وثيقة. تعود فوائد تعاونهما إلى تطوير تقنياتهم وتحسين أدائهم. فالحاسوب يستلم أساليب من الإحصاء لتحليل البيانات واتخاذ القرارات. من جهة أخرى، يستفيد الإحصاء من قوة الحواسيب في معالجة البيانات الكبيرة. وينفذ الحسابات بكفاءة عالية.

4. علاقة الرياضيات بعلم الحاسوب Ma & CS [2,5,8,9,6]

الرياضيات تلعب دوراً حاسماً في علوم الحاسوب، حيث تشكل الركيزة الأساسية التي يُبنى عليها هذا العلم. تمكّن المعرفة الجيدة بالرياضيات المبرمجين وعلماء الحاسوب من فهم كيفية عمل أجهزة الكمبيوتر ووحدات التحكم المنطقية، وتطوير خوارزميات أفضل والعمل على التشفير والحماية.

هناك عدة فروع رياضية ترتبط ارتباطاً وثيقاً بعلوم الحاسوب، منها:

a. الخوارزميات: هي تعليمات تحدد خطوات تنفيذ برامج الحاسوب، وعلم الرياضيات أساس لتطوير الخوارزميات التي تستخدم في البرمجة والحوسبة على سبيل المثال خوارزميات البحث ، الترتيب وتشفير البيانات.

b. الرياضيات المتقطعة: تحليل العلاقات بين الأشياء في صورة تفاصيل محددة، وهي مهمة في البرمجة وعلوم الحاسوب.

c. الاحتمالات: تتيح للبرمجيات التوقع والتعامل مع الحوادث غير المعروفة والبيانات غير الثابتة.

d. حساب التفاضل والتكامل: يستخدم في العديد من مجالات علم الحاسوب، بما في ذلك إنشاء الرسوم البيانية والمحاكاة وتشفير التطبيقات وتصميم الخوارزميات.

e. في عصر البيانات الكبيرة والتعلم الآلي، تعتمد البرمجة بشكل كبير على الإحصاءات والرياضيات الرقمية لتحليل البيانات واستخراج المعلومات القيمة منها.

f. التشفير المستخدم في مجال أمن المعلومات والأمن السيبراني، يعتمد بدرجة كبيرة على التكامل بين مفاهيم وأساليب الرياضيات ونظرية الأعداد في تطوير وفهم تقنيات التشفير والتوقيع الرقمي وأمان الشبكات والاتصالات .

g. التحليل العددي في الرياضيات يعتمد بدرجة أساسية على برامجيّات الحاسوب ومدى تطور أيعازاتها مثل (Matlab, R-package, Python).

أخيراً أصبح العلم تراكمي يكمل بعضه البعض. فالرياضيات هي اساس الاحصاء والإحصاء هي الاساس لعلم البيانات وعلم البيانات هو اساس للذكاء الصناعي الذي يعتمد على برامجيّات الحاسوب المتطورة بأستمرار ومنها Python , R-package وهكذا.

5. علاقة الإحصاء بتقنية المعلومات : ST & IT [12]

نظراً لظهور تخصص علمي جديد في السنوات العشرة الأخيرة ومهم في حقل العمل (تقنية المعلومات IT) لابد أن نبحت علاقته مع أقرب العلوم اليه (الأحصاء ، الحاسوب)، فالإحصاء يستخدم تقنية المعلومات لجمع وتخزين وتحليل وعرض البيانات بطرائق فعالة ودقيقة، وتقنية المعلومات تستفيد من الإحصاء لتطوير وتحسين وتقييم الأنظمة والبرمجيات والخوارزميات المستخدمة في مجالات مختلفة. بعض الأمثلة على تطبيقات الإحصاء في تقنية المعلومات هي:

a. الذكاء الاصطناعي والتعلم الآلي: تستخدم هذه المجالات الإحصاء لتدريب واختبار وتحسين النماذج، والخوارزميات التي تقوم بمحاكاة السلوك البشري والقدرات الذهنية في حل المشكلات واتخاذ القرارات والتعرف على الأنماط والتواصل والإبداع.

b. الأمن السيبراني والتشفير: تستخدم هذه المجالات الإحصاء لتوليد وفك وتحليل الرموز والشفرات التي تستخدم لحماية ونقل المعلومات بشكل آمن وموثوق. كما تستخدم الإحصاء لاكتشاف ومنع ومواجهة الهجمات والتهديدات السيبرانية.

c. التصميم والتطوير والاختبار: تستخدم هذه المجالات الإحصاء لتحديد وتحليل وتحسين متطلبات ومواصفات ووظائف وجودة وأداء الأنظمة والبرمجيات والتطبيقات التي تقدمها تقنية المعلومات للمستخدمين والعملاء.

d. التنقيب عن البيانات والتحليل الكبير: تستخدم هذه المجالات الإحصاء لاستخراج وتنظيف وتنظيم وتحليل وتفسير وعرض البيانات الضخمة والمتنوعة والمعقدة التي تنتجها وتجمعها تقنية المعلومات من مصادر مختلفة. كما تستخدم الإحصاء لاستنباط وتقديم المعرفة والمعلومات والنصائح والتوصيات المفيدة للمستخدمين والعملاء.

6. العلاقة بين علوم الحاسوب وتقنية المعلومات : CS & IT

1.6 علوم الحاسوب هي دراسة المبادئ والتطبيقات الأساسية للحوسبة، بما في ذلك تصميم وتنفيذ واستخدام أجهزة الكمبيوتر والبرامج. يركز هذا المجال على الجانب النظري للحوسبة، ويشمل دراسة المفاهيم الأساسية مثل الخوارزميات والبنية التحتية للبيانات بينما تقنية المعلومات هي دراسة استخدام الكمبيوتر والاتصالات لحل المشكلات والإدارة وإنشاء المعلومات. يركز هذا المجال على الجانب التطبيقي للحوسبة، ويشمل دراسة أنظمة الكمبيوتر والشبكات وقواعد البيانات وتطبيقات البرمجيات.

2.6 ماهو الافضل علوم حاسوب ام تقنيه معلومات؟

الأختيار بين دراسة علوم الحاسوب وتقنية المعلومات يعتمد على الأهداف ومجالات الاهتمام كل منهما له ميزاته واستخداماته الخاصة ويمكن من خلال المقارنة البسيطة بين الاثنين يساعدنا في اتخاذ القرار:

a. علوم الحاسوب (Computer Science):

1. تركيز على البرمجة والخوارزميات: علوم الحاسوب تركز بشكل أساسي على تطوير البرمجيات وتصميم الخوارزميات. إذا كنت مهتمًا بالبرمجة وبناء التطبيقات، فإن علوم الحاسوب قد تكون مناسبة لك.
2. علوم نظام الكمبيوتر: تشمل أيضًا دراسة نظم الكمبيوتر والشبكات وأمان المعلومات. إذا كنت ترغب في فهم كيفية عمل الأنظمة والشبكات وحمايتها، فإن علوم الحاسوب تقدم هذه الفرصة.
3. التعمق في البرمجة: تقدم علوم الحاسوب فرصًا للتعمق في لغات البرمجة وتطوير البرمجيات على مستوى أعمق.

4. المهن: يشمل علماء الكمبيوتر مجموعة واسعة من المهن، بما في ذلك تطوير البرمجيات وهندسة الكمبيوتر وعلوم البيانات.

b. تقنية المعلومات (Information Technology):

1. التركيز على تقنية المعلومات وإدارتها: تقنية المعلومات تركز على استخدام وإدارة التكنولوجيا في العمليات التجارية والمؤسسات. إذا كنت تهتم بإدارة الأنظمة والشبكات والدعم التقني.
2. الاعتماد على التكنولوجيا الحالية: تقنية المعلومات تتطلب فهمًا عميقًا لتكنولوجيا المعلومات الحالية والتوجهات التكنولوجية المستقبلية.

3. **تفاعل مع المستخدمين:** تشمل تقنية المعلومات أيضًا تفاعلًا مع المستخدمين وتقديم الدعم التقني لهم.
4. **المهن :** يشمل متخصصو تكنولوجيا المعلومات مجموعة واسعة من المهن، بما في ذلك هندسة الشبكات وإدارة أنظمة المعلومات ودعم تكنولوجيا المعلومات.
- باختصار، إذا كان الشخص مهتمًا بالبرمجة وتطوير البرمجيات، فعلوم الحاسوب قد تكون الخيار الأفضل. إذا الأهتمام إدارة واستخدام التكنولوجيا لدعم العمليات التجارية، فتقنية المعلومات قد تكون الخيار الأنسب.
7. **الأنشطة العلمية المشتركة [4]**
- من خلال متابعة المواقع الخاصة بالأنشطة العلمية (المؤتمرات والمجلات العلمية) للعلوم الثلاث (الرياضيات، الأحصاء ، الحاسبات) نلاحظ أغلبها تقام بصورة مشتركة (العناوين ، الأهداف ومحاور النشاط) وهذا يدل على التكامل بينهما وأن كل علم منهما يعتمد بصورة أساسية على العلوم الأخرى ويمكن عرض بعض الأمثلة عن ذلك كما يأتي:

Statistical Analysis Conferences 2024/2025/2026

Discipline

3D Bioprinting
3D Printing
3d Ultrasound
4d Ultrasound
Abdominal Hernia
Abdominal Imaging
Abdominal Radiology
Abdominal Ultrasound
Abnormal Child Psychology
Abnormal Psychology
Abortion
Academic Education
Acarology
Access To Care

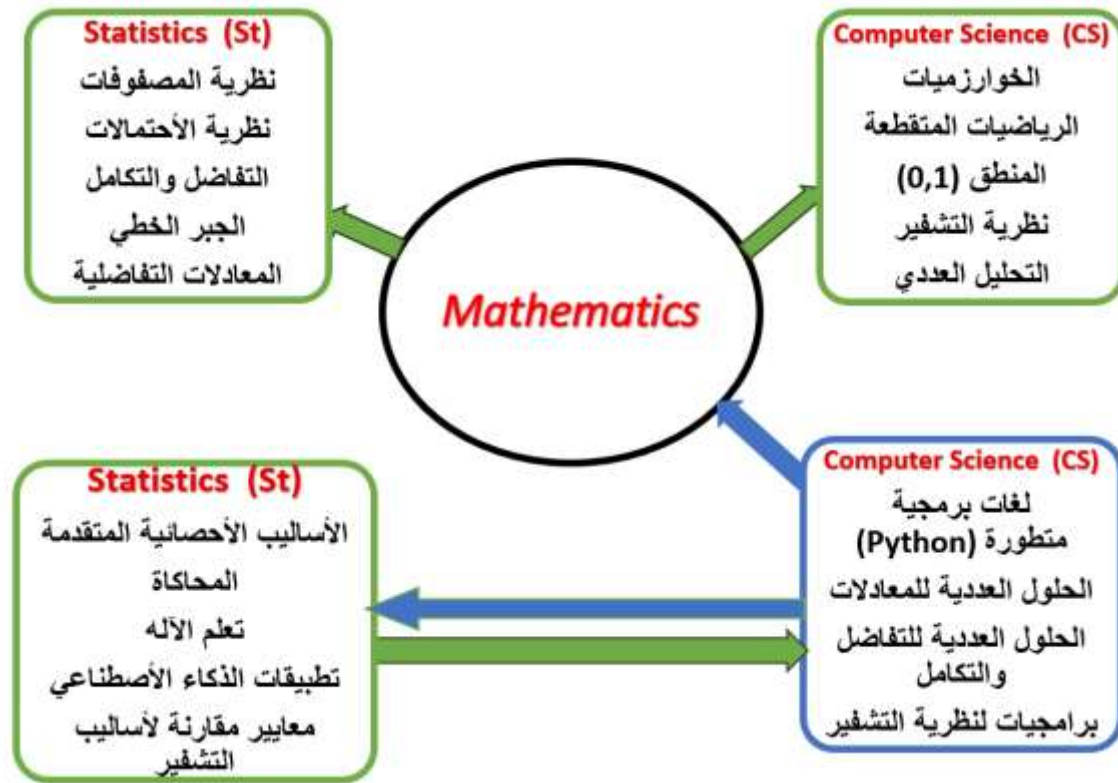
City

October, 2024

Oct 17 International Conference on Statistics, Econometrics and Mathematics (ICSEM) - Quebec City, Canada
Oct 17 International Conference on Mathematical, Computational and Statistical Sciences (ICMCSS) - Istanbul, Turkey
Oct 17 International Conference on Mathematical and Statistical Sciences (ICMSS) - London, United Kingdom
Oct 17 International Conference on Applied Mathematics (ICAM) - Athens, Greece
Oct 17 International Conference on Mathematical and Statistical Sciences (ICMSS) - Quebec City, Canada
Oct 17 International Conference on Mathematical and Statistical Sciences (ICMSS) - Istanbul, Turkey
Oct 17 International Conference on Mathematical and Statistical Sciences (ICMSS) - Kathmandu, Nepal
Oct 17 International Conference on Statistics, Econometrics and Mathematics (ICSEM) - Istanbul, Turkey
Oct 17 International Conference on Mathematical and Statistical Sciences (ICMSS) - Athens, Greece
Oct 17 International Conference on Statistics, Econometrics and Mathematics (ICSEM) - London, United Kingdom
Oct 17 International Conference on Mathematics and Statistical Engineering (ICMSE) - Istanbul, Turkey
Oct 17 International Conference on Computational and Applied Mathematics (ICCAM) - London, United Kingdom
Oct 17 International Conference on Statistics, Econometrics and Mathematics (ICSEM) - Athens, Greece
Oct 17 International Conference on Mathematics and Mathematics Education (ICMME) - Kathmandu, Nepal
Oct 17 International Conference on Mathematical, Computational and Statistical Sciences and Engineering (ICMCSE) - Athens, Greece

شكل رقم (1): صورة من بعض النشاطات العلمية المشتركة بين العلوم الثلاث.

ويمكن عرض مخطط بشكل ملخص يوضح العلاقة بين علم الرياضيات ، الأحصاء وعلم الحاسوب من خلال الشكل الآتي:-



شكل رقم(2): يوضح العلاقة بين العلوم (الرياضيات ، الأحصاء ، الحاسوب)

8. الاستنتاجات

من خلال الاستعراض المرجعي للعلاقة بين العلوم المذكورة في عنوان البحث تم التوصل الى الاستنتاجات الآتية:-

1. وجود علاقة قوية بين العلوم الأساسية الثلاث وأن الرياضيات تعتبر العلم الأم لكافة العلوم الأخرى (الأحصاء والحاسوب) لأعتمادهما على أسس الرياضيات ومفاهيمها .
2. أن التطور الحاصل في علم الأحصاء من خلال النماذج والأساليب الحديثة لم يكتب لها النجاح والبرهنة لولا تطور علوم الحاسوب من الناحية المادية والبرمجية .

3. نظرا" لحاجة علوم الحاسبات للمساهمة في حل المشاكل التي تظهر في الدوال والمعادلات وكل الأسس الرياضياتية ادى الى التعمق في طرح دوال برمجية (Codes) من خلال برامج وتطبيقات حديثة مثل Python, R-packages ساهمت في تطور علوم الحاسوب.
4. من خلال المراجعة هناك علوم أخرى ومنها تقنية المعلومات (تكنولوجيات المعلومات IT) لها علاقة وثيقة مع علم الإحصاء وعلوم الحاسوب وقد شهدت تقنية المعلومات في السنوات الأخيرة تطبقا" مهمة في التعامل مع البيانات وهي مهمة في حقل العمل وهي تجمع بين عدة علوم منها الإحصاء ، العلوم الإدارية المختلفة وعلوم الحاسوب .

References:

1. Burden , R.L.& Faires. (2010), Numerical Analysis ,Cengage Learning.
2. Donald E. K. (1974), Computer Science and Its Relation to Mathematics , the American Mathematical Monthly , Vol. 81, No. 4 (Apr., 1974), pp. 323-343.
3. International Conference on Recent Advances in Mathematics, Statistics and Computer Science 2015,Bihar, India , 29 – 31 May 2015 ,<https://doi.org/10.1142/9651>, July 2016, Pages: 676
4. Irina Arhipova, The Role of Statistical methods in Computer Science and bioinformatics, icots-7, 2006: arhipova, latvia university of agriculture, Latvia
5. Karl Josef Fuchs, (2024), On the Relationship between Mathematics- and Computer Science Education , July 2024 , Teaching Mathematics and Computer Science 22(1):15-34 , DOI:10.5485/TMCS.2024.13412 , Paris Lodron University of Salzburg
6. Ken, B., Cliff, S. ,(2002), Discrete Math in Computer Science, Dept. of Computer Science, Dartmouth College.
7. Rice,J.A. (2006),Mathematical .Statistics and Data Analysis . Cengage Learning.
8. Shreyanjana (16/9/2021), "The Link Between Mathematics and Computer Science Courses", superprof, Retrieved 5/2/2022. Edited.
9. What is the Importance of Mathematics in Computer Science?", geeksforgeeks, 2/8/2019, Retrieved 5/2/2022. Edited.
10. Why computer science students need math ,December 2002 , ACM SIGCSE Bulletin 34(4):57-59 , DOI:10.1145/820127.820166
11. Zill,D.G. & Wright, W.S. ,(2017), Advanced Engineering Mathematics Jones & Bartlett Learning.
12. تطبيق البرامج الاحصائية في الحاسوب – <http://tele-ens.univ-oeb.dz/moodle/course/view.php>