Improving delivery mode forecasting with deep neural network: a time-based convolutional network strategy

Saeed Hamouda[1], Ayman Mohamed[2], Hany A. Elsalamony[3,4]
[1]Department of Networks and Cybersecurity, Faculty of Information Technology, Al-Ahliyya Amman University, Amman, Jordan
[2]Department of Data Science and Artificial Intelligence, Faculty of Information Technology, Al-Ahliyya Amman University, Amman, Jordan
[3]Department of Computer Science, Faculty of Information Technology, Al-Ahliyya Amman University, Amman, Jordan
[4]Department of Mathematics, Faculty of Science, Capital (Helwan) University, Cairo, Egypt

Article Info

Article history:

Received Jun 16, 2025
Revised Oct 11, 2025
Accepted Dec 6, 2025

Keywords:

Caesarean section operation
normal deliveries
Decision tree
Deep learning
Machine learning
Support vector machine

ABSTRACT

The caesarean section is one of the most frequently performed surgical procedures worldwide, with profound implications for maternal and neonatal health. Accurate prediction of delivery mode is essential for guiding clinical decisions, minimizing unnecessary surgical interventions, and improving patient outcomes. This study introduces a deep neural learning technique based on a temporal convolutional neural network (DNLTC) to classify delivery type—caesarean section versus normal vaginal delivery using maternal and obstetric data. The proposed model was evaluated against traditional machine learning (ML) approaches, including artificial neural networks (ANN), support vector machines (SVM), and decision trees (DT). Experimental results show that the DNLTC achieved the highest overall accuracy (85%), surpassing ANN (80%), SVM (68.8%), and DT (65%). TCNN also demonstrated strong clinical reliability, with a sensitivity of 94%, specificity of 91%, and a perfect F1-score of 100%. These findings highlight the advantages of incorporating temporal feature learning into delivery mode prediction, enabling the detection of subtle, sequential patterns that conventional models may overlook. By providing more accurate and robust predictions, the proposed framework can support obstetricians in making timely, evidence-based decisions, ultimately enhancing maternal and newborn health outcomes.

Corresponding Author:

Hany A. Elsalamony
Department of Mathematics, Faculty of Science, Capital (Helwan) University
Cairo, Egypt
Email: h_salamony@science.helwan.edu.eg

1. INTRODUCTION

Accurate prediction of delivery mode, whether a normal vaginal birth or a caesarean section, is a crucial aspect of obstetric decision-making. This choice has significant implications for maternal and neonatal health, influencing complication rates, recovery time, and healthcare costs. While caesarean sections can be lifesaving in emergencies, unnecessary procedures increase the risks of haemorrhage, infection, and long-term reproductive complications. As maternal and obstetric datasets grow in size and complexity, there is a growing opportunity to apply advanced data-driven methods to support clinicians in making these critical decisions with greater precision and consistency [1].

Machine learning (ML) offers powerful tools for extracting predictive insights from clinical and demographic data, enabling earlier intervention and more efficient resource allocation. Traditional statistical approaches often struggle with high-dimensional, non-linear medical data, whereas ML techniques, such as support vector machines (SVM), decision trees (DT), and artificial neural networks (ANN), have shown promise in detecting patterns that are not immediately apparent to human observers. In the context of obstetrics, these algorithms have been applied to predict fetal distress, classify delivery mode, and detect anomalies in cardiotocography (CTG) signals. For example, Huang *et al*. [2] employed DT, ANN, and discriminant analysis (DA) for delivery classification, reporting ANN accuracy of 97.78%, while Ocak [3] used SVM combined with genetic algorithms (GAs) to achieve near-perfect accuracy in distinguishing normal from pathological cases. Other studies have incorporated Naïve Bayes (NB) with feature selection (FS), random forests (RFs), or boosting methods, achieving high accuracies but often omitting critical metrics such as sensitivity and specificity measures that are essential for evaluating clinical reliability.

Despite these advances, three key limitations remain. First, many models rely on static feature sets and fail to capture temporal patterns in medical data, such as variations in maternal vital signs or fetal health indicators over time. Second, imbalanced datasets, common in obstetrics, can bias model performance toward majority classes, leading to unreliable predictions for less frequent but clinically important outcomes. Third, there is a lack of direct comparative evaluation between advanced deep learning models and conventional ML approaches on the same obstetric datasets, making it difficult to determine the most effective method for real-world applications.

This study addresses these gaps by introducing a deep neural learning technique based on a temporal convolutional neural network (DNLTC). Unlike static classifiers, temporal convolutional neural network (TCNN) architectures can model sequential dependencies in clinical variables, improving the detection of subtle risk patterns associated with delivery mode. We benchmark the DNLTC against established methods SVM, DT, and ANN using a publicly available obstetric dataset. Our evaluation includes not only accuracy but also sensitivity, specificity, and F1-score, ensuring a comprehensive assessment of clinical applicability. The contributions of this paper are the development of a TCNN-based classification framework for predicting delivery mode from maternal and obstetric data, and a direct comparison of TCNN performance with widely used ML models (SVM, DT, and ANN) on the same dataset. Then, comprehensive performance analysis, including accuracy, sensitivity, specificity, and F1-score to reflect both predictive power and clinical reliability. Finally, interpretation of findings in the context of reducing unnecessary caesarean sections and improving decision support systems in obstetrics.

The remainder of this paper is organized as follows. Section 2 reviews related work on delivery mode prediction and ML applications in obstetrics. Section 3 introduces the ML techniques, including FS and TCNN architecture. Section 4 describes the dataset and method. Section 5 presents the discussion of the results and comparative performance analysis. Section 6 introduces the experimental results. Finally, section 7 concludes the paper by summarizing contributions and highlighting the potential for TCNN integration into real-time decision support systems.

2. RELATED WORKS

Numerous papers regarding the caesarean section procedure are available in the literature. Huang *et al*. [2] reviewed DT, ANN, and DA as classifiers within comparative investigations. The ANN classifier yielded a total accuracy of 97.78%. The other two classifiers, the DT and DA, came in second and third place with accuracy rates of 86.36% and 82.1%, respectively. As noted, the performance estimates do not include sensitivity and specificity components, making accuracy alone a questionable metric, particularly for binary classifiers. For evaluations where the datasets are heavily weighted in one class and the prior probabilities differ significantly, accuracy presents this problem. Similar research by Ocak [3] examined SVM and GA classifiers for normal and pathological instances, reporting accuracy rates of 99.3% and 100%, respectively. The same results were noted in [4], [5]. As before, these studies did provide sensitivity and specificity data. Predicting food quality through the detection of certain compounds using sensors was also done with KNN, DT, and LDA [6]. Alam *et al.* [7] used several ML techniques on radiographic images for bone fracture detection.

The researchers focused on the methods and problems encountered during caesarean section in [8], [9]. One of the main problems with a caesarean section is the complication of overwhelming hemorrhage. Blood loss, together with the associated studies, can be found in [10]. Comprehensive details on the prevention of fatal injury while performing caesarean delivery, classification, and risk factors are provided in [11]. For advanced studies and developments, one can look into [12], [13]. Work by Menai *et al*. [14] incorporated a NB classifier with four FS methods: mutual information, correlation-based, ReliefF, and information gain. It was found that having the NB classifier along with features created by ReliefF gave the best results for classifying foetal state, achieving 93.97% accuracy, 91.58% sensitivity, and 95.79% specificity.

In the work by Karabulut and Ibrikci [15] shows that the contribution of AdaBoost ensemble is to C4.5 DT and accuracy is improved up to 95.01%. Spilka *et al.* [16] using the CTG-UHB dataset reported CTG-UHB dataset, with 72 and 78% sensitivity and specificity values, with LCA based RF classifiers. With the same dataset, attempted to identify hypoxia using the C4.5 DT, NB, and SVM, the SVM was shown to have the highest results, 73.4% sensitivity and 76.3% specificity.

3. MACHINE LEARNING TECHNIQUES

ML applications have made great strides in the last few years in the area of clinical diagnostics and have shown promise for other clinical applications as well, such as disease prevention, diagnosis, prognosis, drug discovery, and clinical trial design [17], [18]. Generally speaking, supervised learning algorithms have shown the most promise in the area of clinical diagnostics to date and are the most commonly used. However, unsupervised learning algorithms and reinforcement learning are also available for use and are better suited for some problems in the clinical setting as shown in Figure 1 and described in the Glossary.

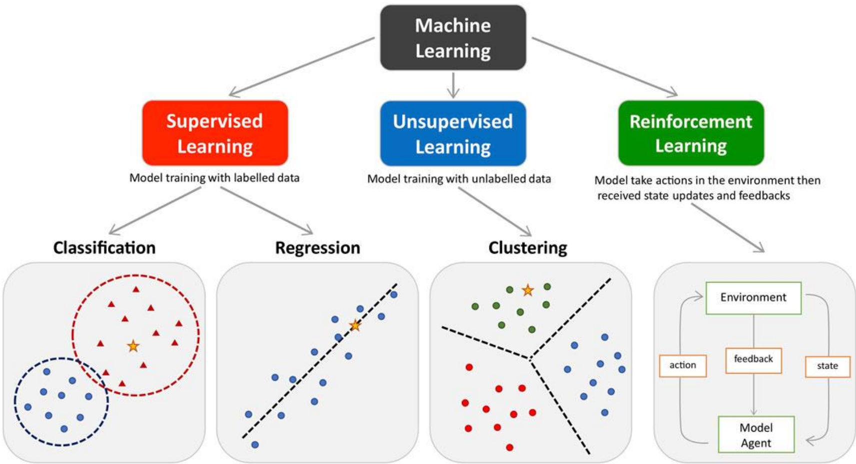


Figure 1. Main types of ML: supervised (classification, regression), unsupervised (clustering), and reinforcement learning

3.1. Deep neural learning technique classifier

A technique from deep neural learning (DNLTC) uses temporal convolution and builds on nearest neighbor classification. Its dynamic component allows the method to adapt to local trends in the data. This method enhances static distance measures by considering the developing patterns in time series data, and it consequently makes a better prediction of the class labels.

3.2. Support vector machine

The development of SVMs has been carried out by Cortes and Vapnik [19], Cristianini and Scholkopf [20], and Joachims [21], and these methods are gaining a lot of attention thanks to various nice features and what seems to be very promising experimental performance. In the ML community, SVM is a well-known technique and celebrated by many for being a state-of-the-art method that has performed very well over the past decade. When you look at the application domain of SVMs, they cover a lot of interesting ground, as you can see in the next paragraph, and they occupy a prominent role in a lot of different ML research areas [22].

3.3. Neural network

A popular ML method is ANNs with back propagation (BP) [23], which has numerous advantages, like better approximation capabilities. Still, it has certain drawbacks, like the selection of the number of hidden layer neurons, slow convergence, and an imprecise learning rate, among others. ANNs are different from conventional methodologies in that they can be trained by examples to solve the problem, rather than being told what to do by a fixed algorithm [24]-[26]. For more details, interested readers can consult references [27]-[30].

3.3. Decision tree

The reason that DTs are popular is that they are simple to understand and easy to explain. If you have an uncertain situation that you want to decide about a DT can give you a strategic answer [31]. It can also handle problems with nonlinear relationships quite well [32].

4. MATERIALS AND METHODS

Data used in our approach is obtained from the UC irvine ML repository [33]. Having cesarean results conducted on 80 pregnant females with most severe delivery troubleshooting at the hospital department familiar, and that gathered and used for ML classifiers that are ANN, SVM, and DT testing and training. The cesarean section and normal vaginal deliveries are classified with the help of deep learning by several means. One is to scan the medical imaging data to determine the mode of delivery using TCNN. The models are trained on extensive medical record databases that contain data about the type of delivery performed and then used to predict the type of delivery to be performed on future patients.

This article categorizes an obstetric database of caesarean sections into four levels of urgency. These outpatient categories, recommended by the UK National Confidential Enquiry into Patient Outcome and Death (NCEPOD), the Royal College of Obstetricians and Gynaecologists (RCOG), and the Royal College of Anesthetists (RCA), are now well established in UK practice. 80 case data are employed for training and for assessing ML classifiers, specifically, ANN, SVM, and DT, for the task of predicting uterine rupture. The data instances have set a caesarean section operation as 'yes' and class (1). In contrast, those cases having the final verdict that a caesarean section surgery should not be performed are classified (0) and assigned the value 'no'. The technique used most for initial labor fetal monitoring is CTG. Clinical decisions are usually based on the visual examination of the CTG traces. The poor human interpretation of them has a reason, though. A large body of research has shown that obstetricians have little agreement with each other when it comes to even the basic interpretation of CTG. In essence, poor human CTG interpretation is the reason why some very large interval studies have led to the kind of very poor consensus we have today. That interval study poor consensus then leads to poor obstetrical outcomes, like unnecessary cesarean sections, which is a big CTG interpretation cost factor problem. In this study, we use stochastic gradient descent (SGD) to minimize the loss function. We also utilize the dropout regularization technique. This guarantees that, when a specific training sample is used, the activity of each neuron in the network is suppressed with probability P during forward propagation. For input neurons, this coefficient is normally 0.2, while for hidden neurons, it is 0.5. Dropout allows an ensemble of an exponentially large number of models to be averaged, which reduces overfitting and improves generalization. We modify backpropagation using momentum and learning rate annealing so that previous iterations can affect the current version of the model. In particular, we define a velocity vector, v, to change the updates.

The model's loss function is minimized using a standard (SGD) optimization procedure. We apply dropout regularization during training, where neurons are randomly deactivated with predefined probabilities, using rates of 0.2 for input layers and 0.5 for hidden layers [32]. By preventing co-adaptation of neurons, dropout typically leads to improved generalization performance, as it effectively approximates the averaging of an ensemble of sub-networks [34].

5. RESULTS AND DISCUSSIONS

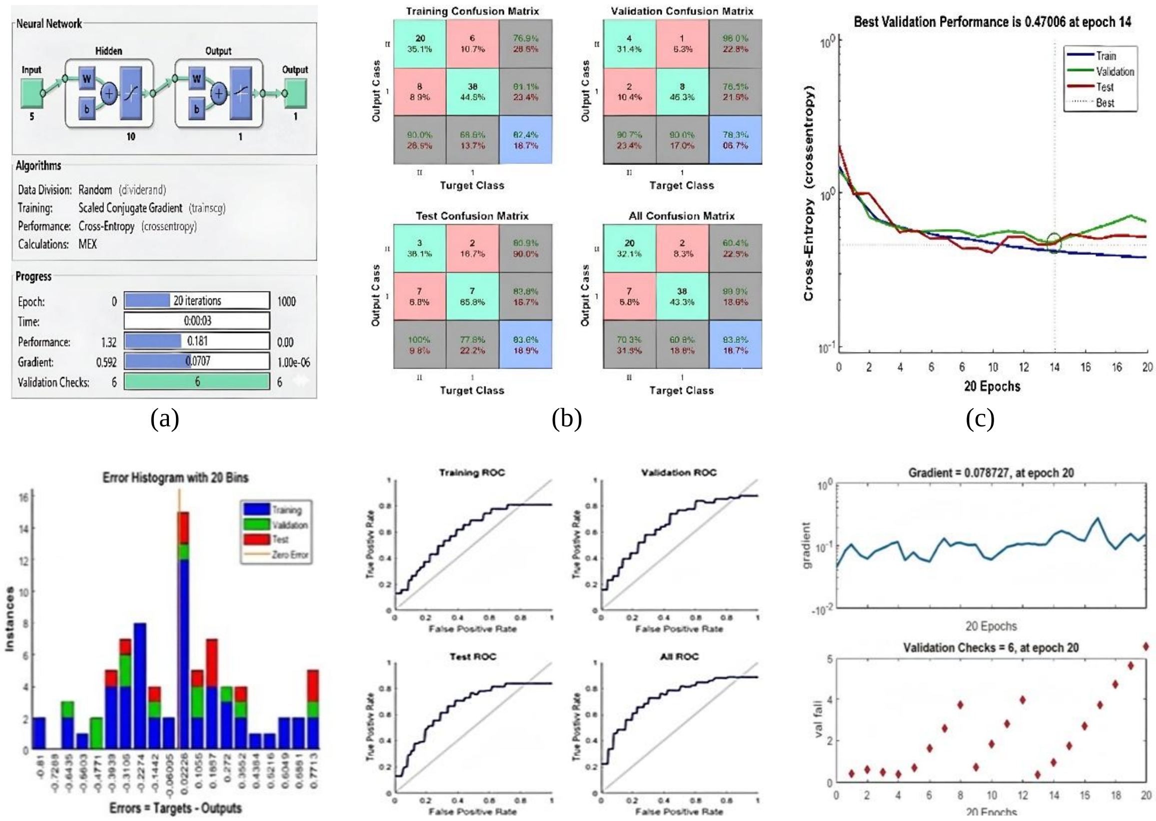
The following are the results and overall performance of four common classification methods: ANN, SVM, DT, and a specific type of neural network, the TCNN. Almost ALL the results you will see in this paper, and the associated algorithms, were created on a Windows 10 operating system running MATLAB R2017a and an Intel Core i7@ 2.6 GHz with 16 GB RAM. A caesarean section can be divided into four categories. Category (1) represents the immediate threats to the mother's life or that of the foetus, whereas (2) accounts for mothers' or foetuses' compromise that is not immediately life-threatening. In (3), early delivery is required with no compromise to mothers or foetuses, while (4) shows the delivery time suitable for the woman and staff (elective). The attributes in the dataset are the input variables and are named as follows: age of instance, number of pregnancies, time of delivery, blood pressure, and heart status. These five attributes are applied to 80 instances. Table 1 represents these attributes and their kinds.

Table 1. The attributes and their kinds

Attributes	Age	No. of pregnant	Delivery time	Blood pressure	Heart status	Caesarean
Kinds	Numerical	Numerical	Time, premature, latecomer	Low, normal, high	Inept, apt	Yes, no
Number represented	17:40	1:4	0, 1, 2	0, 1, 2	1, 0	1, 0

5.1. Artificial neural networks

One of the key strengths of ANNs is their ability to handle multidimensional and non-linear associations between variables and outcomes. This is particularly valuable in obstetrics, where the decision for a C-section is influenced by a dynamic and complex interplay of factors like maternal age, BMI, gestational age, and fetal health. Figure 2 shows NN training results by presenting Figure 2(a) the neural network feed forward, Figure 2(b) the confusion matrix, Figure 2(c) the validation performance, Figure 2(d) the histogram of errors, Figure 2(e) the receiver operating characteristic (ROC) for training model (validation and testing), and Figure 2(f) display the gradient and validation checks. In Figure 2(a), the network is performing at 0.03 seconds of the training data time consumed in 20 iterations out of 1000 epochs with a performance of 0.381 and a gradient of 0.0787 within the interval [0.592, 1.00e-06]. The error indicating samples which are not classified (unexhibitable samples) are 19.6, 25, and 16.66 for train, validation, and test sets, respectively. Also, minimum cross-entropy results in correct classification, which states that for training, validation, and testing give 0.506, 0.7287, and 0.7299, respectively.



(a) (b) (c)

(d) (e) (f)

Figure 2. All results of applied ANN on caesarean section dataset; (a) the ANN network, (b) the confusion matrix, (c) the validation performance, (d) histogram of errors, (e) the ROC curve, and (f) the gradient and validation checks

All the confusion matrices pertaining to the training, validation, and testing processes of the network NN in Figures 2(c)–(f), the manuscript presents a comprehensive explanation of the validation performance, error histogram, ROC curves, and gradient/validation checks, ensuring full clarity and alignment with editorial requirements. In the up left quarter, the training confusion matrix holds two green examples among the five cells featuring the right answer and corresponding proportion by the count of instances with the correct classification. For further granularity, out of the 20 instances (samples), 20 are correctly classified, as they will not result in doing the cesarean section operation with a TN. The analysis focuses on the predictive accuracy of a classification model concerning caesarean section operations. Out of 80 total instances analyzed, the model correctly classified 35.7% of cases as either undergoing or not undergoing the procedure. Specifically, 25 instances were accurately identified as those that would undergo a caesarean, reflecting a 44.6% success rate. However, there were instances of misclassification, with 10.7% wrongly identified as candidates for surgery when they were not, and 8.9% of those who undergo the procedure misclassified as not. The data reveals a high level of accuracy in predictions overall. For cases classified as not undergoing the operation, 76.9% were predicted correctly, while 83.3% of those predicted to undergo the caesarean were accurate. The training confusion matrix indicates that 80.4% of predictions were correct, with additional validation showing 75% accuracy and 83.3% during testing. The model's performance metrics indicate 80 true positives (TP), 27 true negatives (TN), 7 false positives (FP), and 9 false negatives (FN). In conclusion, the classification model demonstrates a robust predictive capability, achieving an overall accuracy of 80%. The sensitivity and specificity measures also indicate reliable performance, with values of 80.43% and 79.41%, respectively. These results suggest that while the model is effective, there remains room for improvement in reducing misclassifications to enhance overall diagnostic accuracy.

The study focuses on evaluating the performance of different DT models in a classification task. It highlights the validation performance, reaching its optimal value of 0.47006 at epoch 14, indicating potential overfitting during earlier epochs. The training, validation, and test curves displayed distinct patterns, with significant improvements noted in the test curve. Figure 2(e) illustrates the training errors and the ROC curves, suggesting effective classification as the curves are positioned closer to the upper left corner.

5.2. Support vector machine

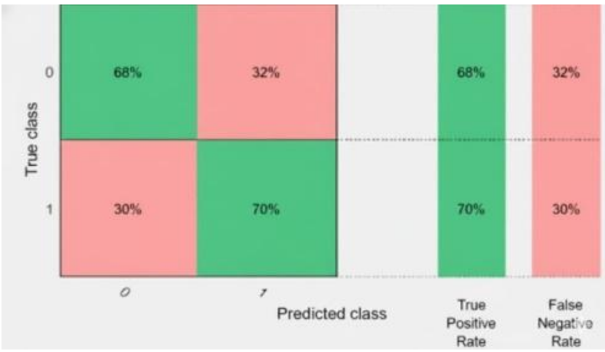
SVMs often perform well because they are good at handling complex, non-linear relationships in the data. However, their performance can be sensitive to the choice of kernel and hyperparameters, and they may be less effective when the number of features is much greater than the number of samples. The study evaluates various SVM models applied to cesarean data, focusing on their performance metrics such as accuracy, sensitivity, and specificity. The findings, summarized in Table 2, detail the outcomes of different models, along with their training times, kernel scales, and number of observations. A 5-fold cross-validation approach was employed to ensure the reliability of the results, with specific attention to the target attribute outlined in Table 1. Among the models assessed, the Quadratic SVM achieved the highest accuracy at 68.8% and specificity of 67.65%, requiring 1.9141 seconds for training with an automatic kernel scale. In contrast, the coarse Gaussian model recorded a remarkable sensitivity of 100%, albeit with a specificity of 0%, and had a training time of 1.7186 seconds. The linear model performed the worst in terms of accuracy, achieving only 56.3%, while the fine Gaussian model exhibited the lowest sensitivity at 76.09%. Notably, both fine and medium Gaussian models displayed identical accuracies of 65%.

Table 2. Different kinds of SVM models

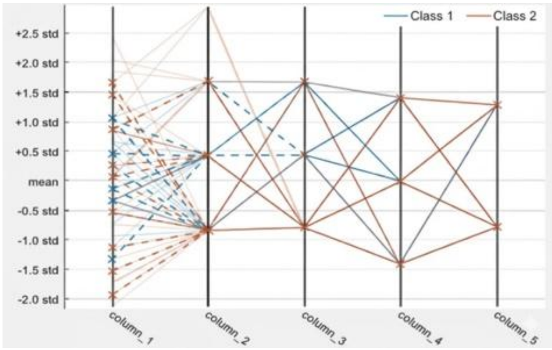
SVM	Accuracy (%)	Sensitivity (%)	Specificity (%)	Training time	Kernel scale	Number of observations			
						TP	TN	FP	FN
Quadratic SVM	68.8	69.57	67.65	19.141	Auto	32	23	11	14
Cubic SVM	67.5	78.26	52.94	17.148	Auto	36	18	16	10
Fine Gaussian	65	76.09	50	13.478	0.56	35	17	17	11
Medium Gaussian	65	82.6	41.18	12.364	2.2	38	14	20	8
Coarse Gaussian	57.5	100	0	17.186	8.9	46	0	34	0
Linear SVM	56.3	89.13	11.76	20.081	Auto	41	4	30	5

In conclusion, the results indicate that while the Quadratic SVM model excels in accuracy and specificity, the coarse Gaussian model stands out in sensitivity, highlighting the trade-offs between different performance metrics. The findings underscore the importance of selecting appropriate SVM models based on the specific objectives of data classification tasks, particularly in medical datasets where both precision and recall are critical.

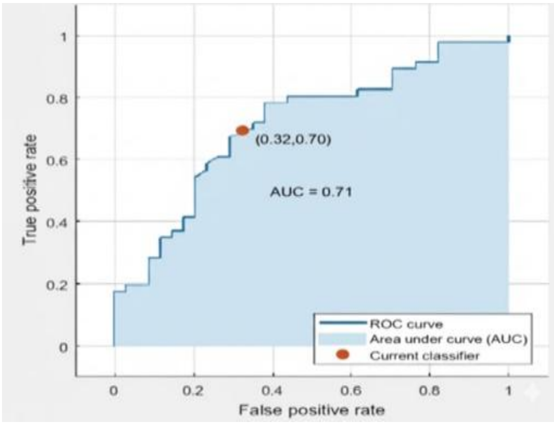
The green cell on the bottom row in Figure 3(a) contains all instances that will do the operation of caesarean sections, all having a true class. In the columns provided, it is shown that 70% of the instances have been correctly classified as will do the operation of caesarean section, thus having a 70% TP rate for correctly classified value in this class. Also, in the above green cell, 68% of the instances have been correctly classified as will not do the operation of caesarean section. Thus, so has a 68% TN rate for classified value in the correct of class as green column TP rate. The other instances, which are in the same row, remain misclassified. They have been marked in red color cells: 30% of the personnel are incorrectly classified as will do the operation, and, in turn, 32% incorrectly as will not do the operation of caesarean section. Thus, termed FN rate for incorrectly classified class values, red cells. Figure 3(b) presents a parallel coordinates graph for understanding relationships between features and classifying the useful attributes by using the separating classes with visualized training data and misclassified values plotted in dashed lines, whereas classified ones in lines, (0)'s in orange lines, and (1)'s classes in color blue. Figure 3(c) gives the ROC curve.



(a)



(b)



(c)

Figure 3. Quadratic SVM model, the trained attributes, and ROC graph; (a) Quadratic SVM model confusion matrix, (b) attributes of Quadratic SVM model, and (c) ROC of Quadratic SVM model

5.3. Decision tree

The simplicity of DTs can be both a strength and a weakness. They are prone to overfitting, especially with complex datasets, which means they may not generalize well to new data. Their performance can be highly dependent on the structure of the tree, and even a small change in the data can lead to a very different tree. The results, summarized in Table 3, reveal that the complex tree achieved the highest accuracy at 65%, with 27 TP and 25 TN, while recording 9 FP and 19 FN. The medium tree excelled in sensitivity, reaching 70.72%, whereas the simple tree achieved the highest specificity at 76.47%.

Table 3. Different kinds of DTs

DT	Accuracy (%)	Sensitivity (%)	Specificity (%)	Training time	Number of observations			
					TP	TN	FP	FN
Complex tree	65	58.7	73.53	0.6037	27	25	9	19
Simple tree	62.5	52.17	76.47	0.35641	24	26	8	22
Medium tree	61.3	70.72	53.49	0.38932	26	23	20	11

These findings demonstrate the varying strengths of each tree model in handling the classification of pregnant women based on the dataset. In conclusion, the analysis underscores the effectiveness of DT algorithms in classification tasks, with each model exhibiting unique advantages. While the complex tree provided the best accuracy, the medium tree's high sensitivity and the simple tree's strong specificity highlight the importance of selecting the appropriate model based on specific classification needs. This study contributes valuable insights into the application of DTs in healthcare-related data analysis, emphasizing the need for careful evaluation of model performance across multiple metrics.

The analysis of DTs applied to caesarean section operation data reveals varying levels of accuracy and performance among different tree complexities. The complex DT achieved the highest accuracy at 65% with a training time of 0.6037 seconds, utilizing a maximum of 100 splits and 10 surrogate splits based on deviance reduction. Conversely, the simple DT, while faster at 0.35641 seconds, achieved 62.5% accuracy and demonstrated the highest specificity at 76.4%, but had lower sensitivity at 52.17%. The medium DT performed the least effectively with an accuracy of 61.3% and specificity of 53.49%, taking 0.38932 seconds for training. Figures 4(a)–(c) displays the tree structures; Figures 4(d)–(f) displays the confusion matrices; and Figures 4(g)–(i) present ROC curves for complex, simple, and medium trees respectively.

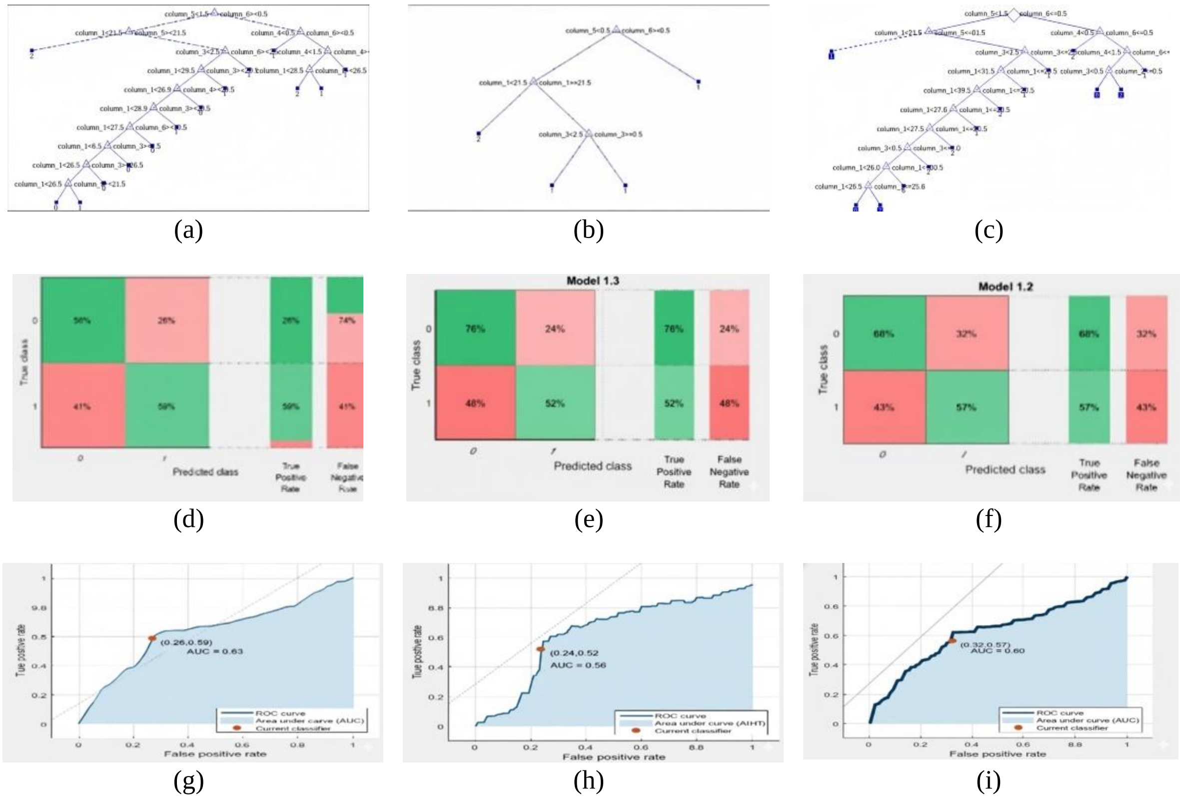


Figure 4. The kinds of DT constructions and performances; (a) the complex DT, (b) the simple DT, (c) the medium DT, (d) complex tree confusion matrix, (e) simple tree confusion matrix, (f) medium tree confusion matrix, (g) ROC of complex tree, (h) ROC of simple tree, and (i) ROC of medium tree

In further detail, the construction and pruning of the DTs vary, with the complex and medium trees pruned to 7 levels, while the simple tree was pruned to 3 levels. Each tree utilized a maximum of 10 surrogate decision splits, which aids in enhancing model performance. The confusion matrix for the complex DT illustrates its effectiveness, showing that 59% of the instances were accurately predicted for undergoing a caesarean section (TP rate), while 74% were correctly identified as not needing the operation (TN rate). In conclusion, while the complex DT outperformed the others in accuracy, the simple DT's higher specificity highlights the different strengths of each model. The results emphasize the importance of model selection based on the specific needs of classification tasks, as accuracy, sensitivity, and specificity can vary significantly. This analysis serves as a useful reference for future DT applications in medical contexts.

The text discusses the performance evaluation of DT models for classification tasks, specifically focusing on the accuracy rates of simple, medium, and complex DTs. The confusion matrices reveal that the simple DT achieved a TP rate of 52% and a TN rate of 76%, while misclassifications were recorded at 48% for FP and 24% for FN. The medium DT showed slight improvements with TP and TN rates of 57% and 68%, respectively, and corresponding FP and FN rates of 43% and 32%. In contrast, the complex DT had a TP rate of 59% and an FP rate of 26%, indicating a better classification ability compared to the simpler models.

5.4. Temporal convolutional neural network

TCNNs and other deep learning models are very good at automatically learning complex features from raw data, which is a significant advantage over traditional methods where features need to be manually engineered. This makes them particularly effective for time-series data like CTG traces. However, these models can be computationally expensive to train, require large amounts of data to perform well, and their "black box" nature can make it difficult to understand the reasoning behind their predictions.

The TCNN architecture comprises three convolutional layers followed by max pooling layers and two fully connected layers, ultimately classifying the data into two output nodes. The training process of TCNN employs SGD with backpropagation to optimize performance, measured through accuracy, precision, recall, and F1-scores on a testing set. By leveraging a TCNN, medical decision-making can potentially be enhanced, resulting in better outcomes for patients through data-driven assessments of delivery methods.

The discussed model employs a binary cross-entropy loss function alongside the Adam optimizer, featuring specific hyperparameters such as a learning rate of 0.0001 and beta values of 0.9 and 0.999. Figure 5 shows that the training process involves 500 epochs with a batch size of 32, while 10% of the training data is set aside for validation. The evaluation metrics used to assess model performance include accuracy and Logloss, ensuring a comprehensive analysis of the model's effectiveness. Key findings indicate that the model’s performance on the validation set peaked with a weight parameter (W) set to 200, achieving the highest area under curve (AUC) and the lowest Logloss values. The Logloss metric converged around 0.50 after completing the 500 epochs, suggesting stable performance without significant overfitting. These results underscore the model's capability to generalize well to unseen data while maintaining robustness throughout the training process.

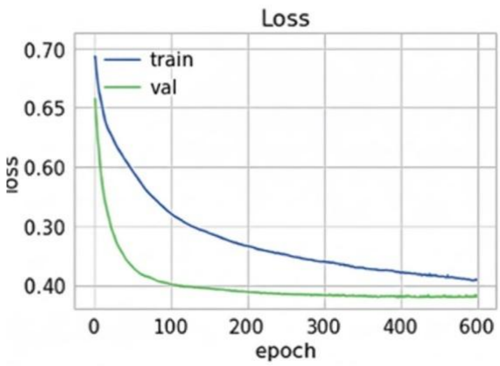


Figure 5. TCNN training and validation for NCEPOD dataset

6. EXPERIMENTAL RESULTS

These results indicate that while many artificial intelligence (AI) models are effective, the best-performing model can depend heavily on the specific characteristics of the dataset. The ultimate goal of these models is to serve as a clinical decision support system, providing doctors with valuable, real-time information to help them make better decisions for their patients:

- TCNN models excel with time-series data and often achieve the highest accuracy and other metrics, especially when combined with data pre-processing techniques.
- ANNs can achieve a high level of accuracy in predicting C-sections. While specific results vary widely based on the dataset, features used, and network architecture.
- SVM is a strong performer, especially with well-structured, non-time-series data, and can achieve very high accuracy with proper tuning.
- DT provide a good baseline and are highly interpretable, but they generally have lower accuracy compared to more advanced models and can be prone to overfitting.

7. CONCLUSION

This study demonstrates the potential of deep learning, particularly TCNNs, in improving the prediction of delivery mode in obstetrics. Using maternal and obstetric data, the proposed deep neural learning technique (DNLTC) achieved higher predictive accuracy than conventional ML approaches, outperforming ANN, SVM, and DT models. Notably, the TCNN delivered strong sensitivity and specificity, underscoring its reliability for clinical decision support. Compared with traditional classifiers, the temporal learning capability of TCNN enables it to capture sequential patterns and subtle variations in patient data that static models may overlook. This strength is particularly valuable in obstetric decision-making, where early recognition of risk factors can guide timely interventions and reduce the likelihood of unnecessary caesarean sections. The comparative evaluation provided in this study also offers clarity on model performance trade-offs, which is often missing in existing literature.

The findings have practical implications for developing intelligent decision support systems in maternity care. By integrating TCNN-based models into clinical workflows, healthcare providers could enhance the consistency and accuracy of delivery mode predictions, leading to better maternal and neonatal outcomes. In addition, the comprehensive use of performance metrics beyond accuracy ensures that such models are assessed in a clinically meaningful way.

Future research should focus on validating the proposed approach with larger, more diverse datasets, exploring multi-modal inputs such as real-time physiological signals and imaging data, and addressing dataset imbalance through advanced resampling or cost-sensitive learning techniques. Expanding the scope to include other obstetric outcomes could further increase the impact of this work on maternal healthcare. In summary, this research contributes both a methodological advancement through the application of TCNN to delivery mode prediction, and a comparative framework for evaluating ML and DL models in clinical settings. These results suggest that deep temporal architectures can play a significant role in advancing precision medicine in obstetrics, bridging the gap between data analytics and real-world clinical decision-making.

FUNDING INFORMATION

Authors state that no funding was received for this research. This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Saeed Hamouda	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓		✓
Ayman Mohamed		✓	✓	✓		✓	✓	✓	✓	✓	✓		✓	✓
Hany A. Elsalamony	✓		✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this study, as no new data were created or analyzed. The dataset used in this research is publicly available at the UCI Machine Learning Repository and can be accessed through the corresponding reference in the manuscript. The authors confirm that the data supporting the findings are available at the UCI Machine Learning Repository (Ref. [33]). Derived data supporting the findings of this study are available from the corresponding author (H.A.E.) upon reasonable request.

REFERENCES

[1] B. Milovic and M. Milovic, "Prediction and Decision Making in Health Care using Data Mining," *International Journal of Public Health Science (IJPHS)*, vol. 1, no. 12, pp. 69-78, 2012.

[2] Huang, Mei-Ling, and Yung-Yan Hsu, "Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network," *Journal of Biomedical Science and Engineering*, vol. 5, no. 9 , pp. 526-533, 2022, doi: 10.11591/IJPHS.V1I2.1380.

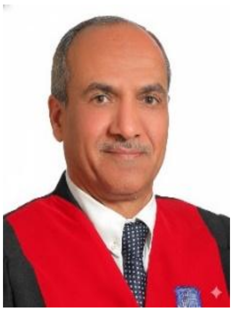
[3] H. Ocak, "A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being," *Journal of Medical Systems*, vol. 37, no. 2, 2013, doi: 10.1007/s10916-012-9913-4.

[4] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer*, vol. 29, no. 3, pp. 31–44, Mar. 1996, doi: 10.1109/2.485891.

[5] G. Cybenko, "Neural networks in computational science and engineering," *IEEE Computational Science & Engineering*, vol. 3,

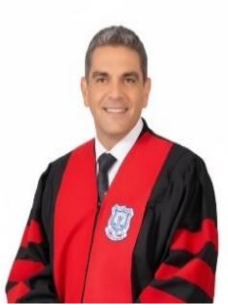
no. 1, pp. 36–42, 1996, doi: 10.1109/99.486759.
[6] M. A. Al Sharaiah *et al*., "Deep Learning Model Based on E-Noses Sensors in Food Production Controlling System," *Tuijin Jishu/Journal of Propulsion Technology*, vol. 45, no. 3, 2024.
[7] A. Alam *et al.,* "Novel transfer learning based bone fracture detection using radiographic images," *BMC Medical Imaging*, vol. 25, no. 1, p.p. 1-16, Jan. 2025, doi: 10.1186/s12880-024-01546-4.
[8] A. Simm and P. Ramoutar, "Caesarean section: Techniques and complications," *Current Obstetrics and Gynaecology*, vol. 15, no. 2, pp. 80–86, Apr. 2005, doi: 10.1016/j.curobgyn.2004.12.003.
[9] A. Simm and D. Mathew, "Caesarean section: techniques and complications," *Obstetrics, Gynaecology and Reproductive Medicine*, vol. 18, no. 4, pp. 93–98, Apr. 2008, doi: 10.1016/j.ogrm.2008.01.007.
[10] E. A. Bonney and J. E. Myers, "Caesarean section: Techniques and complications," *Obstetrics, Gynaecology and Reproductive Medicine*, vol. 21, no. 4, pp. 97–102, Apr. 2011, doi: 10.1016/j.ogrm.2010.12.002.
[11] S. Shah-Becker, B. S. Oberman, and J. G. May, "Scalping of a newborn: Complication during cesarean section," *International Journal of Pediatric Otorhinolaryngology Extra*, vol. 10, no. 3, pp. 53–55, Sep. 2015, doi: 10.1016/j.pedex.2015.04.002.
[12] I. Hardy and S. Rousseau, "Captive uterus syndrome: An unrecognized complication of cesarean sections?," *Medical Hypotheses*, vol. 122, pp. 98–102, Jan. 2019, doi: 10.1016/j.mehy.2018.10.010.
[13] S. Gobillot *et al.*, "Obstetric outcomes associated with induction of labour after caesarean section," *Journal of Gynecology Obstetrics and Human Reproduction*, vol. 47, no. 10, pp. 539–543, Dec. 2018, doi: 10.1016/j.jogoh.2018.09.006.
[14] M. El B. Menai, F. J. Mohder, and F. Al-mutairi, "Influence of Feature Selection on Naïve Bayes Classifier for Recognizing Patterns in Cardiotocograms," *Journal of Medical and Bioengineering*, vol. 2, no. 1, pp. 66-70, 2013. doi: 10.12720/jomb.2.1.66-70.
[15] E. M. Karabulut and T. Ibrikci, "Analysis of cardiotocogram data for fetal distress determination by decision tree based adaptive boosting approach," *Journal of Computer and Communications*, vol. 2, no. 9, pp. 32-37, 2014, doi: 10.4236/jcc.2014.29005.
[16] J. Spilka, G. Georgoulas, P. Karvelis, V. Chudáček, C. D. Stylios, and L. Lhotská, "Discriminating normal from "abnormal" pregnancy cases using an automated fhr evaluation method," *Hellenic Conference on Artificial Intelligence,* Cham: Springer International Publishing, 2014, vol. 8445, pp. 521–531, doi: 10.1007/978-3-319-07064-3_45.
[17] I. S. Stafford, M. Kellermann, E. Mossotto, R. M. Beattie, B. D. MacArthur, and S. Ennis, "A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases," *NPJ Digital Medicine,* vol. 3, 2020, doi: 10.1038/s41746-020-0229-3.
[18] S. J. MacEachern and N. D. Forkert, "Machine learning for precision medicine," *Genome*, vol. 64, no. 4, pp. 416-425, 2021, doi: 10.1139/gen-2020-0131.
[19] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning,* vol. 20, no. 3, pp. 273-297, 1995, doi: 10.1007/BF00994018.
[20] N. Cristianini and B. Scholkopf, "Support vector machines and kernel methods: the new generation of learning machines," *Ai Magazine*, vol. 23, no. 3, pp. 31-31, 2002, doi: 10.1609/aimag.v23i3.1655.
[21] T. Joachims, *Making large-scale svm learning practical*, Advances in Kernel Methods-Support Vector Learning, MIT Press, Cambridge, MA, USA, pp. 169-184, 1999.
[22] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, Jun. 1998, doi: 10.1023/A:1009715923555.
[23] R. P. Lippmann, "Pattern Classification Using Neural Networks," *IEEE Communications Magazine*, vol. 27, no. 11, pp. 47–50, Nov. 1989, doi: 10.1109/35.41401.
[24] W. G. Baxt, "Use of an Artificial Neural Network for Data Analysis in Clinical Decision-Making: The Diagnosis of Acute Coronary Occlusion," *Neural Computation*, vol. 2, no. 4, pp. 480–489, Dec. 1990, doi: 10.1162/neco.1990.2.4.480.
[25] S. Kamruzzaman, Md. A. Hamid, and A. J. Sarkar, "ERANN: An Algorithm to Extract Symbolic Rules from Trained Artificial Neural Networks," *IETE Journal of Research*, vol. 58, no. 2, pp. 138-154, 2012.
[26] H. Temurtas, N. Yumusak, and F. Temurtas, "A comparative study on diabetes disease diagnosis using neural networks," *Expert Systems with Applications*, vol. 36, no. 4, pp. 8610–8615, May 2009, doi: 10.1016/j.eswa.2008.10.032.
[27] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, Mar. 1986, doi: 10.1007/bf00116251.
[28] R. E. Banfield, L. O. Hall, K. W. Bowyer, and W. P. Kegelmeyer, "A comparison of decision tree ensemble creation techniques," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 173–180, Jan. 2007, doi: 10.1109/TPAMI.2007.250609.
[29] P. H. Swain and H. Hauska, "The decision tree classifier: Design and potential," in *IEEE Transactions on Geoscience Electronics*, vol. 15, no. 3, pp. 142-147, Jul. 1977, doi: 10.1109/TGE.1977.6498972.
[30] C. Sehra, "Decision Trees Explained Easily," Medium, Jan. 19, 2018. [Online]. Available: https://chirag-sehra.medium.com/decision-trees-explained-easily-28f23241248. (Accessed on May 10, 2025).
[31] R. Shafique, A. S. Al-Shamayleh, S. K. Posa, A. Ishaq, F. Rustam, and G. S. Choi, "Advancing ovarian cancer outcomes with CTGAN-enhanced hybrid machine learning approach, *Knowledge-Based Systems*, vol. 328, p. 114206, 2025, doi: 10.1016/j.knosys.2025.114206.
[32] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
[33] M. Amin and A. Ali, "Caesarian Section Classification Dataset," UCI Machine Learning Repository, 2018, doi: 10.24432/C5N59X.
[34] H. A. Elsalamony, "Detection of anaemia disease in human red blood cells using cell signature, neural networks and SVM," *Multimedia Tools and Applications*, vol. 77, no. 12, pp. 15047–15074, Jun. 2018, doi: 10.1007/s11042-017-5088-9.

BIOGRAPHIES OF AUTHORS



Saeed Hamouda received a Bachelor's degree in Systems and Computer Engineering from Al-Azhar University, Egypt, a Master's degree in Systems and Computer Engineering from Al-Azhar University, Egypt, and Ph.D. in Systems and Computer Engineering from Al-Azhar University, Egypt. He was the Head of the Computer Science Department at the College of Computer Science, Nahda University. Currently, he is Assistant Professor in Faculty of Information Technology, Department of Networks and Cyber Security, Al-Ahliyya Amman University, Amman, Jordan. His research interests are intelligent systems and cybersecurity. He can be contacted at email: s.hamouda@ammanu.edu.jo.

Ayman Mohamed received a Bachelor of Science in Scientific Computing from Zagazig University, Egypt, a master's degree in information systems from the Institute of Computer Science and Artificial Intelligence from Helwan University, Egypt, and a Ph.D. in Information Systems from Helwan University, Egypt, with a dissertation titled "Intelligent Decision Support System Approach to Mitigate Climate Change Effects". He was the Vice Dean for Student Affairs at Badya University. Currently, he is Assistant Professor in Faculty of Information Technology, Department of Data Science and Artificial Intelligence, Al-Ahliyya Amman University, Amman, Jordan. His research interests are decision support system, expert systems, and artificial intelligence. He works as a Researcher at Agricultural Research Center (ARC). He can be contacted at email: a.mohamed@ammanu.edu.jo or ayman.mohamed@arc.sci.eg.



Dr. Hany A. Elsalamony is an Associate Professor of Computer Science at Al-Ahliyya Amman University, Jordan, and an Associate Professor of Computer Science and Mathematics at Helwan University, Egypt. He earned his B.Sc. (1995), M.Sc. (2001), and Ph.D. (2006) in computer science and mathematics from Helwan University, where his doctoral work on wavelet-based human motion detection and face recognition earned the Best Ph.D. Thesis Award. Over more than two decades, he has held leadership roles including Vice Dean, Head of the Department of Computer Science, and Head of Research Projects Unit across institutions in Egypt and colleges in Saudi Arabia. His research spans computer vision, image processing, machine learning, and mathematical modeling, with high-impact publications and an active role as a reviewer for international journals and IEEE conferences. He can be contacted at email: h.salamony@ammanu.edu.jo.