Bulletin of Electrical Engineering and Informatics
Vol. 12, No. 2, April 2023, pp. 1219~1225
ISSN: 2302-9285, DOI: 10.11591/eei.v12i2.4180

Smart evaluation for deep learning model: churn prediction as a product case study

Esam Mohamed Elgohary[1], Mohamed Galal[2], Ahmed Mosa[3], Ghada Atef Elshabrawy[4]
[1]Information Systems, Institute of National Planning, CLIP Project Manager, Cairo, Egypt
[2]Department of Predictive Analytics, National Bank of Egypt, Cairo, Egypt
[3]Mini-Max Software Solution, Mansoura, Egypt
[4]R&D and Training Manager, Mini-Max Software Solution, Mansoura, Egypt

Article Info

Article history:

Received Jun 1, 2022
Revised Aug 13, 2022
Accepted Oct 18, 2022

Keywords:

Churn prediction
Customer relationship management
Data mining
Deep learning
Predictive models

ABSTRACT

Customer churn prediction recently is one of the vital issues that confronts diverse business industries to sustain the customers base and profits. On the other hand, data scientists employ gigantic customer data to automate the data modelling process to offer these models as a generally portable service. This research has two main contributions: deep learning customer churn prediction model and smart evaluation prediction model service. So, this service harnesses any customer data to automate building, evaluation, and deployment of the churn prediction model. The research consists of three main parts. Firstly, it illustrates the dataset labelling which annotates customers data into churn or non-churn. Secondly, the deep learning churn prediction framework using convolutional neural network (CNN) algorithm. Finally, a case study is presented to show how churn prediction service is automatically trained and generated based on real customer data, where CNN parameters are adapted to achieve the most reliable performance in line with customers' behavior. The applied case study achieves accuracy 0.77, area under the curve (AUC) 0.84 and f1 score 0.83.

Corresponding Author:

Esam Elgohary
Information systems, Institute of National Planning, CLIP Project Manager
Al-Tayaran st., Cairo, Egypt
Email: esam.elgohary@inp.edu.eg

1. INTRODUCTION

Machine learning models were used for predicting in different fields [1]-[4]. Nowadays, diverse business markets reach a congestion state and face a brutal competition between different service providers. This competition arises due to the market saturation of abundant service providers and the products' offers diversity. Herein, churn prediction is a business use case, which applies various data mining techniques to detect the customers who are likely to cancel their subscription to a special service [5], [6]. Customer behavior changes in line with the defined business use case. Customizing each prediction model costs redundant time and effort for each case [7]. Herein, data scientists automate the data modelling process, so it generalizes the modelling process and offers it as a service [8]. This research comes as part of implementing a customer relationship management (CRM) system called customer loyalty intelligent personalization (CLIP). CLIP is a smart, machine learning based personalized customer advisory system. CLIP can serve different kinds of business applications. It aims to assist E-commerce and retail businesses to retain their profits and their customer base. This paper proposes a framework for automating customer churn prediction with respect to the business use case. The paper sections are: section 2 outlined a literature review, section 3 illustrated the customer labelling dataset, section 4 showed an overview on the implemented framework, and section 5 showed a case study on client's real data.

2. BACKGROUND

Churn prediction has been studied in various researches, with different aims. Authors focused on comparing different techniques and approaches in particular domains. This section outlined some of the recent ongoing researches, as following: Research by Castanedo *et al.* [9], performed churn prediction by leveraging deep learning architectures image classification. Firstly, supervised learning was performed on over 6 million customers using deep convolutional neural networks (CNNs), which achieved an area under the curve (AUC) of 0.778 on the test dataset. The study's main weakness is the extremely scarce user-user input data, which grows significantly when long-term user interactions are taken into account. Therefore, an input data architecture may encode these long-term connections among users in order to better predict long-term interactions in a telecommunication dataset

Another research by Wangperawong *et al.* [10], in order to do churn prediction, customer temporal behavioral data was represented as images using deep learning architectures popular in image classification. Deep CNNs were used for supervised learning on labelled data from over 6 million users, and they produced an AUC of 0.743. However, no more than 12 temporal features were employed for each customer. To increase the effectiveness of the input photographs, more features can be added.

Research by Ismail *et al.* [11], in one of the top telecommunications firms in Malaysia, a multilayer perceptron (MLP) neural network approach has been presented to forecast customer turnover. Its outcome shown that in prediction tasks, neural networks were superior to statistical models (91.28% prediction accuracy). There were only 78 churners and 58 nonchurners in the training set, but there were 13 churners and 10 nonchurners in the testing set. This data set is incredibly little, and it cannot possibly be used to forecast churn in the telecom sector. Additionally, the dataset utilised is private, thus performance comparison is not possible.

Research by Tariq *et al.* [12], the suggested model employs a 2-D (CNN; a technique of deep learning). The suggested model features a layered design with two distinct phases: a layer for data import and preprocessing, and another layer for 2-D CNN. A parallel environment is also employed to process the data using the Apache Spark distributed and parallel framework. Telco customer churn is used to extract Kaggle training data. An accuracy rating of 0.963 out of 1 was assigned to the suggested model. Additionally, there is a very small loss during training and validation (0.004). The true-positive and true-negative values are 95% and 94%, respectively, according to the results of the confusion matrices. They simply reported the performance data; they did not compare their models with any other models; however, the false-negative is only 5% and the false-positive is only 6%, which is effective.

Cenggoro *et al.* [13] employed a vector embedding model to estimate loss for a telecom dataset of 3,333 users without contrasting the suggested model with any other models. The model's accuracy and F1-score were the only metrics provided, and they show that the model does a good job of differentiating between churning and non-churning clients. Zhou *et al*. [14] proposed a model based on long short term memory network (LSTM) and CNN which has cross-layer connections between the LSTM layers and the convolution layers. This model learns the latent sequential information and captures important local features from time series features. Experimental results on the real-life dataset showed that the proposed model performed better than other comparison models.

Zhong and Li [15] proposed the CNN-based predictive model to detect churn signals from transcript data of phone calls. Experimental results showed that when sufficient training data was provided with our text annotation method, their CNN-based predictive model generated state-of-the-art performance in churn prediction. Finally, Pirmohammadi and Mast [5] proposed multi-layer perceptron ANN with 8 neurons in a hidden layer has applied and the best performance of this network appears in epoch 10. Then, the structural model of the network was added. On the other hand, a Regression test has been used in order to predict customer churn by SPSS. The best performance of ANN occurred in epoch 10. The statistical performance of ANN model, in order to classify output and target value, SME and RSME are computed which are equal to 0.15599 and 0.39495 respectively.

This research proposed an automated solution for the customer churn prediction model. It applied a deep learning algorithm to predict future customers' churn rates based on real client data. The proposed solution applied a CNN algorithm to build the churn prediction model. It automatically labelled the customer data, then divided it into training, validation, and test sets. Following that, the CNN parameters are adapted to achieve the best prediction model based on customer behavior with respect to the given business case. The implemented software is portable and customizable for generic e-commerce and retail business cases.

First essential step in churn prediction is to assess each customer's behavior. The customers' data isn't labelled or classified before as churned or not, however it can be inferred from their previous purchase transactions. This research proposed a methodology that illustrates how to infer customers' behavior if it is a possible churn or not. The main churn configurable are calculated from the real customer data, which are: purchase frequency bypass times, reduction average purchase percentage, and average reduction purchase times. Figure 1 shows the pseudo-code to annotate customer's data to churn or not churn. Firstly, it fetches customer's first and last visit dates. Secondly, it calculates the number of customer purchases times and their average purchase values. Thirdly, it starts labeling customers churned or not based on the above obtained values. After that, data is divided into training, validation, and testing sets for further processes.

LABELING CUSTOMER DATASET

```
Customers Labeling - Pseudo Code

1. Fetch client last purchase date

2. Fetch customer first visit date (first purchase date)

3. Fetch customer last visit date (last purchase date)

4. Calculate purchase frequency
        Days of (last purchase date – first purchase date) / customer's purchases count

5. Calculate average purchase value
        Customer's total Purchases / customer's purchases count

6. Determine customer's number of purchases
        If number of purchases is 1 THEN label customer as NOT APPLICABLE

7. Determine churn by frequency bypass times
        If (client last purchase date – customer last purchase date) / customer purchase frequency >=
        frequency bypass times THEN label customer as CHURNED

8. Determine churn by purchase reduces for last n times (n reduces times from configurable section)
        nFrequencies = Get customer purchase for last n frequencies
        Foreach frequency in nFrequencies
                Calculate percentage between frequency purchases and average purchases
        If all percentage calculation <= reduce percentage
                THEN label customer as CHURNED

9. Otherwise, label customer as NOT CHURNED
```

Figure 1. Customer labelling pseudocode

3. CUSTOMER CHURN PREDICTION FRAMEWORK

The CNN machine learning algorithm is one of the most famous deep learning algorithms, whose main power is feature engineering without need for domain expertise [5]-[16]. In this research, the CNN algorithm is applied to build the customer churn prediction model. The CNN hyperparameters such as weight constraint, dropout rate, filter numbers, dense neuron number, Kernel size, batch size, and momentum, are initialized randomly. Then, these hyperparameters are repeatedly changed to fit the built CNN model on the customer data. The output model is considered a customized churn prediction model, which can be deployed in a specific business case.

Figures 2 and 3 show the workflow to transform automatically customer raw data into a customized churn prediction model. The input data represents the customer behavior, which includes the feature columns and the actual label if the customer is considered as a churner or non-churner. The output of that workflow is a prediction model. Figure 2 shows the first basic processes to prepare data before data modelling, which are preprocessing, feature engineering, and data splitting. The input in Figure 2 is the raw customer data and the output will be three main organized datasets: training, testing, and validation datasets. The training data size is adjusted to be 60% of full data and 20% is for both validation and testing to report the model performance. Figure 3 shows the details of data auto-modelling processes that work on getting the best fitted prediction model with respect to the provided customer data. The main datasets which are previously formed, are the input for the main three auto-data modelling processes which are training, validating, and testing processes. In the auto-training process, the CNN hyperparameters form a list of combinations. These lists are automatically applied to build and train the CNN algorithm to generate different prediction models. In the auto-validating process, the validation dataset evaluates each generated CNN model in the training process; the highly accurate model is saved. In the auto-testing process, the testing dataset evaluates the accepted CNN model and reports the evaluation metrics and saves for further predicting unseen customer data.

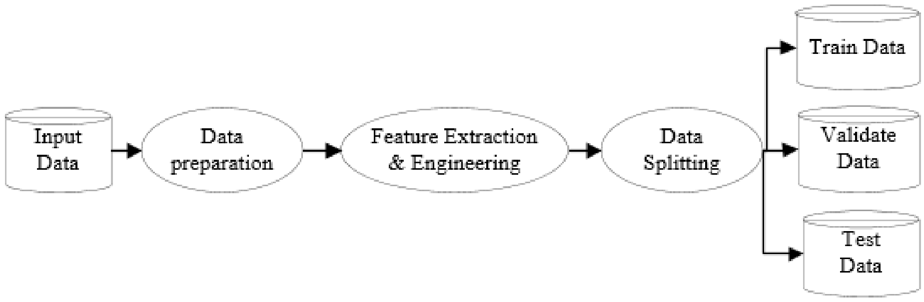


Figure 2. Churn prediction model framework

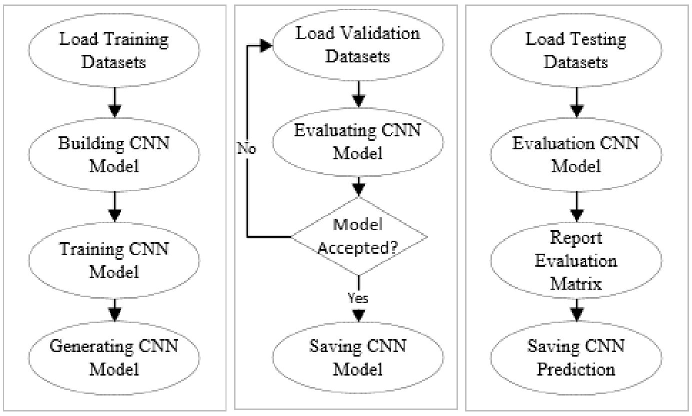


Figure 3. Data modelling processes

A case study presents a sample of real customer data which is used to build a customized churn prediction model. This section's purpose is to show how the prediction model is auto-trained and auto-evaluated to achieve the most recommended model. The four main evaluation numbers to evaluate any supervised machine learning algorithm are: true positive (TP), true negative (TN), false positive (FP), and false negative (FN) [17], and these numbers are considered basics for various evaluation equations like accuracy and F1 score as shown in (1) and (2) respectively. In the churn prediction case, TP represents the number of truly predicted customers as a churner. TN shows the number of truly predicted customers as non-churners. FP displays the number of customers who are non-churners but the predictive algorithm has labelled them as churners. FN represents the number of customers who are churners but the predictive model has labelled them as non-churners.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{2}$$

Any CNN has a collection of hyperparameters as aforementioned [16]-[23]. Each hyperparameter has a preferred value range and impact on the built model performance. Herein, these hyperparameters could form different lists of combinations of various values. Each combination is applied to build, train and evaluate the CNN using the abovementioned equations. In this case study, the applied data size is 476, 119, and 149 for training, validation and testing respectively. Table 1 shows 20 out of 768 lists of combinations of CNN hyperparameters to view their impact on both model training and validation performance. The successful accepted model attained accuracy 0.78 in training and 0.77 in testing, and attained f1 score 0.85 in training and 0.83 in testing.

On the other hand, receiver operating characteristics (ROC) and area under the ROC (AUC) [24], [25] are other evaluation measurements which evaluate the model performance based on TP and FP rates. Figure 4 shows two evaluation graphs for the successful fit model generated in this case study. The right graph displays the ROC and AUC curve which shows highly reliability 0.84 in predicting unseen data, and the graph on the left F1 score for both training and testing.

Table 1. CNN hyperparameters experiments

Weight constraint	Learn rate	Momentum	Dropout rate	CNN filters	Kernel size	Dense neuron	Batch size	Acc. train	F1 train	Acc. Val.	F1 Val.
4	0.001	0.1	0.4	32	3	64	10	0.78	0.85	0.77	0.85
4	0.001	0.1	0.4	32	3	128	10	0.77	0.84	0.74	0.83
4	0.001	0.1	0.4	32	3	256	10	0.75	0.84	0.74	0.84
4	0.001	0.1	0.4	32	4	64	10	0.77	0.84	0.78	0.86
4	0.001	0.1	0.4	32	4	128	10	0.76	0.84	0.73	0.83
4	0.001	0.1	0.4	32	4	256	10	0.78	0.86	0.74	0.84
4	0.001	0.1	0.4	64	3	64	10	0.79	0.85	0.78	0.87
4	0.001	0.1	0.4	64	3	128	10	0.80	0.86	0.78	0.87
4	0.001	0.1	0.4	64	3	256	10	0.75	0.84	0.73	0.84
4	0.001	0.1	0.4	64	4	64	10	0.78	0.85	0.81	0.88
4	0.001	0.1	0.4	64	4	128	10	0.75	0.84	0.73	0.83
4	0.001	0.1	0.4	64	4	256	10	0.75	0.84	0.74	0.84
4	0.001	0.1	0.4	128	3	64	10	0.77	0.85	0.77	0.85
4	0.001	0.1	0.4	128	3	128	10	0.78	0.85	0.78	0.86
4	0.001	0.1	0.4	128	3	256	10	0.75	0.84	0.74	0.84
4	0.001	0.1	0.4	128	4	64	10	0.78	0.85	0.79	0.87
4	0.001	0.1	0.4	128	4	128	10	0.81	0.87	0.79	0.87
4	0.001	0.1	0.4	128	4	256	10	0.76	0.85	0.74	0.84
4	0.001	0.1	0.4	256	3	64	10	0.77	0.84	0.77	0.85

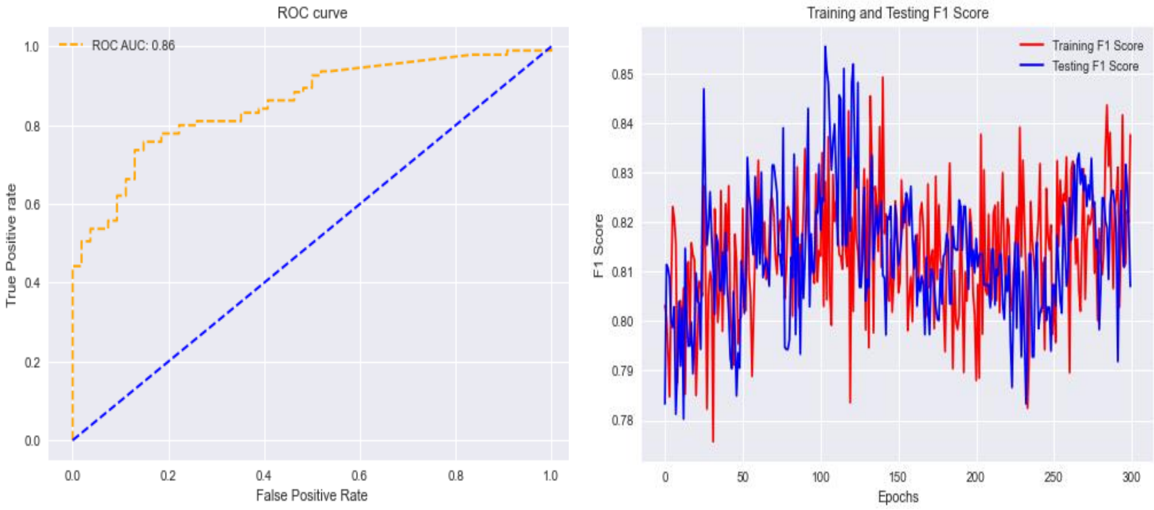


Figure 4. Case study evaluation graphs

4. CONCLUSION

In the thriving technological era, the markets are overloaded with various services providers, which escalates competition between companies to preserve their customer bases and financial gains. Churn prediction is a problem which has intrigued various researchers and business leaders recently. On the other hand, customer data modeling in each business case to generate a churn prediction model consumes too much time and effort. So, this research proposed an automated customer churn prediction service using the CNN algorithm. It facilitates generation of a deep learning churn prediction model for each business case based on their customer behavior. A case study is presented to show the automatic adaptation of CNN hyperparameters until a decision made to select the best fit model. This case study shows reliable AUC measurement reached 0.84. This research can contribute to automatically predicting and evaluating customer churn rates in both e-commerce and retail business applications.

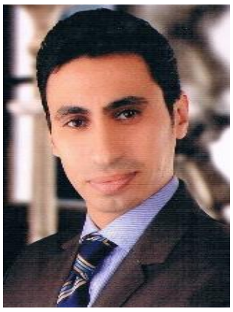
REFERENCES

[1] U. Sa'adah, M. Y. Rochayani, and A. B. Astuti, "Knowledge discovery from gene expression dataset using bagging lasso decision tree," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 2, pp. 1151-1159, Feb. 2021, doi: 10.11591/ijeecs.v21.i2.pp1151-1159.

[2] A. H. Yassir, A. A. Mohammed, A. A.-J. Alkhazraji, M. E. Hameed, M. S. Talib, and M. F. Ali, "Sentimental classification analysis of polarity multi-view textual data using data mining techniques," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 5, pp. 5526-5534, Oct. 2020, doi: 10.11591/ijece.v10i5.pp5526-5534.

[3] N. N. Alabid and Z. D. Katheeth, "Sentiment analysis of Twitter posts related to the COVID-19 vaccines," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 3, pp. 1727-1734, Dec. 2021, doi: 10.11591/ijeecs.v24.i3.pp1727-1734.
[4] K. Yothapakdee, S. Charoenkhum, and T. Boonnuk, "Improving the efficiency of machine learning models for predicting blood glucose levels and diabetes risk," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 1, pp. 555-562, Jul. 2022, doi: 10.11591/ijeecs.v27.i1.pp555-562.
[5] G. Pirmohammadi and M. Mast, "Comparison of Artificial Neural Network and SPSS Model in Predicting Customers Churn of Iran's Insurance Industry," *International Journal of Computer Applications*, vol. 176, no. 32, pp. 14–21, Jun. 2020, doi: 10.5120/ijca2020920345.
[6] M. A. Muslim and Y. Dasril, "Company bankruptcy prediction framework based on the most influential features using XGBoost and stacking ensemble learning," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 6, pp. 5549-5557, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5549-5557.
[7] A. H. Alsaeedi, A. H. Aljanabi, M. E. Manna, and A. L. Albukhnefis, "A proactive metaheuristic model for optimizing weights of artificial neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 2, pp. 976-984, Nov. 2020, doi: 10.11591/ijeecs.v20.i2.pp976-984.
[8] D. L. García, À. Nebot, and A. Vellido, "Intelligent data analysis approaches to churn as a business problem: a survey," *Knowledge and Information Systems*, vol. 51, no. 3, pp. 719–774, Sep. 2016, doi: 10.1007/s10115-016-0995-z.
[9] F. Castanedo, G. Valverde, J. Zaratiegui, and A. Vazquez, "Using deep learning to predict customer churn in a mobile telecommunication network," *Erişim tarihi*, vol. 30, 2014.
[10] A. Wangperawong, C. Brun, O. Laudy, and R. Pavasuthipaisit, "Churn analysis using deep convolutional neural networks and autoencoders," *arXiv*, Apr. 18, 2016, doi: 10.48550/arXiv.1604.05377.
[11] M. R. Ismail, M. K. Awang, M. N. A. Rahman, and M. Makhtar, "A Multi-Layer Perceptron Approach for Customer Churn Prediction," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 10, no. 7, pp. 213–222, Jul. 2015, doi: 10.14257/ijmue.2015.10.7.22.
[12] M. U. Tariq, M. Babar, M. Poulin, and A. S. Khattak, "Distributed model for customer churn prediction using convolutional neural network," *Journal of Modelling in Management*, vol. 17, no. 3, pp. 853–863, May 2021, doi: 10.1108/jm2-01-2021-0032.
[13] T. W. Cenggoro, R. A. Wirastari, E. Rudianto, M. I. Mohadi, D. Ratj, and B. Pardamean, "Deep Learning as a Vector Embedding Model for Customer Churn," *Procedia Computer Science*, vol. 179, pp. 624–631, 2021, doi: 10.1016/j.procs.2021.01.048.
[14] J. Zhou, J. Yan, L. Yang, M. Wang, and P. Xia, "Customer Churn Prediction Model Based on LSTM and CNN in Music Streaming," *DEStech Transactions on Engineering and Technology Research*, no. aemce, May 2019, doi: 10.12783/dtetr/aemce2019/29520.
[15] J. Zhong and W. Li, "Predicting Customer Churn in the Telecommunication Industry by Analyzing Phone Call Transcripts with Convolutional Neural Networks," in *Proceedings of the 2019 3rd International Conference on Innovation in Artificial Intelligence - ICIAI 2019*, 2019, doi: 10.1145/3319921.3319937.
[16] A. Mishra and U. S. Reddy, "A Novel Approach for Churn Prediction Using Deep Learning," in *2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Dec. 2017, doi: 10.1109/iccic.2017.8524551.
[17] H. Hoover, "Predicting & Preventing Banking Customer Churn By Unlocking Big Data," 2017. [Online], available at: https://docplayer.net/3363428-Predicting-preventing-banking-customer-churn-by-unlocking-big-data.html.
[18] Y. Xu and H. Zhang, "Convergence of deep convolutional neural networks," *Neural Networks*, vol. 153, pp. 553–563, Sep. 2022, doi: 10.1016/j.neunet.2022.06.031.
[19] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, Mar. 2021, doi: 10.1186/s40537-021-00444-8.
[20] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Jun. 2018, doi: 10.1007/s13244-018-0639-9.
[21] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–21, 2021, doi: 10.1109/tnnls.2021.3084827.
[22] J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, May 2018, doi: 10.1016/j.patcog.2017.10.013.
[23] D.-X. Zhou, "Theory of deep convolutional neural networks: Downsampling," *Neural Networks*, vol. 124, pp. 319–327, Apr. 2020, doi: 10.1016/j.neunet.2020.01.018.
[24] J. A. Swets, "ROC Analysis Applied to the Evaluation of Medical Imaging Techniques," *Investigative Radiology*, vol. 14, no. 2, pp. 109–121, Mar. 1979, doi: 10.1097/00004424-197903000-00002.
[25] J. A. Hanley, "Receiver operating characteristic (ROC) methodology: the state of the art," *Crit Rev Diagn Imaging*, vol. 29, no. 3, pp. 307–335, 1989.

BIOGRAPHIES OF AUTHORS



Esam Mohamed Elgohary Associate Professor of Information systems and Digital Transformation, Institute of National Planning. He works as a Consultant at Institute of National Planning, Egypt, and consultant at Saudi Commission for Tourism and Antiquities for Information systems applications. He has attended more than 12 workshops and events in USA and Europe for improving academic research and success stories in e-government. He executes many projects in e-government and for private sectors. He can be contacted at email: esam.elgohary@inp.edu.eg.

Mohamed Galal Department of Predictive Analytics, National Bank of Egypt. Department of Computer Science, Ain Shams University Cairo, Egypt. He can be contacted at email: mhdgalal@yahoo.com.

Ahmed Mosa General Manager at Mini-Max software, along more than 17 years of experience in Management sector attached with software engineering sector. He contributed in several IT communities and projrcts. He can be contacted at email: amosa@minimax-soft.com.

Ghada Atef Elshabrawy Assistant professor at New Damietta High Institute of Hotels and Tourism Studies, and Research and Development Manager at Mini-Max software. Her researches focus on the new technologies in marketing and management. She has contributed in many technology development projects. She can be contacted at email: Ghadaelshabrawy85@gmail.com.