

The Effect of the Visualization Type of Data Received from the X Platform on Manipulation Detection with Deep Learning

Hafzullah İş^{1*}

¹Department of Computer Engineering, Batman University, Batman, Turkey

*(hafzullah.is@batman.edu.tr)

Abstract – Deep learning can provide successful results in detecting manipulations made with bot accounts on the X (Twitter) platform and preventing disinformation. As a method for applying deep learning algorithms, numerical values in the data set can be visualized and classified by making sense of them. The methodology used in the data visualization process and the structure of the shape created are critical for deep learning and classification performance. In this article, the effect of the format used to visualize the data in the dataset to be applied deep learning on the classification performance has been demonstrated with more than 400 experimental studies. From visualization methods; The effect of deep learning applied to images created with formats such as area graph, qrcode, spectrogram graph, color map, histogram and distribution on the classification performance with Convolutional Neural Network algorithms has been demonstrated comparatively. As a result of experimental studies, it was observed that while the highest performance was achieved with 98.67% depending on the visual and algorithm used, the performance decreased to 23.70% when different visuals and algorithms were used. In this regard, it has been determined that the methodology applied in revealing the profile confidence index, the effect of the variables that make up the data set and the algorithms used are very important, but the data visualization method used in creating the figure is important. has a critical impact on performance.

Keywords – X, Social Network Analysis, Deep Learning, Transfer Learning, Data Visualization, CNN.

I. INTRODUCTION

In today's world, as a result of developments in information technologies, there is a transformation in people's interactions with each other. As a result of the internet being accessible all over the world and the increase in mobile phone usage, this transformation has begun to occur faster. While the subject of this transformation is humans, the intermediary object is social networks [1]. As Web 2.0 and the Internet pave the way for multifaceted interaction, people have found another medium to express themselves. The versatile interaction opportunities in this medium have been quite attractive. For this reason, diversity in social networks has begun to increase and the number of people using these networks has increased exponentially.

According to We Are Social's "Digital in 2020" report [1], digital, mobile and social media have become an indispensable part of daily life for people all over the world. Social media has become the source of digital culture. With digital culture, communication platforms, socialization areas and topics have changed. This change has caused very effective transformations in many areas, including economic, political and social. The opportunity to communicate and interact independently of geographical borders and without language, religion and culture limitations has made the social networks that the new generation grew up with an integral part of daily life. More than 4.5 billion people currently use the internet, while social media users have surpassed 3.8 billion.

However, the shadow of data misuse remains, with almost half of internet users using ad blockers[1]. Social networks'

user profile structures are important due to the size and diversity of the area that the intense user base affects politically, socially and economically. Social media manipulation by bot accounts makes social network analysis more important.

Figure 1 shows the closeness of bot account interaction to human interaction, which is critical in showing the destructiveness of manipulation in social networks.

Analyzing the profile structures of manipulated users in social networks is valuable in terms of determining their characters [3]. In this regard, understanding the categorical profile structures of users is valuable in preventing manipulation and disinformation in social networks [4]. There are many academic studies addressing the impact and destructiveness of manipulation in social networks. Ferrera's work predicting this danger in social networks has attracted a lot of attention in the literature [5]. In the "Manipulation Policy in Social Media-2020" report of the University of Amsterdam, it was stated that these destructive activities on social media have increased exponentially [6]. For this reason, determining the authenticity of social network user profiles as well as revealing their quality is valuable in preventing destructive manipulation activities.

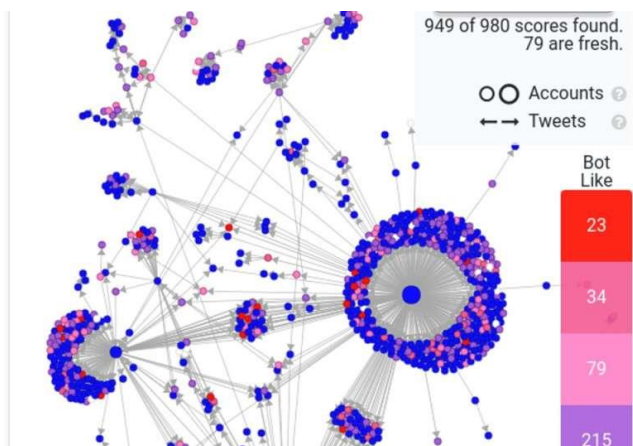


Fig. 1 Similarity Study of Bot Interaction in Social Networks to Human Interaction[2]

The article discusses the methodological approaches of the method, which was developed to detect and prevent the source of manipulation and disinformation in social media, specifically X (Twitter), and whose performance was tested on an up-to-date and comprehensive data set. In the applied methodology, deep learning performance was determined depending on the format of the images used. The data set consists of data that measures a user's interaction with static and dynamic aspects in multiple ways, using 10 different metrics. These metrics; It consists of modularity, reciprocity, centrality, diameter, density, likes, retweets, followers and follows, as well as the total number of tweets sent. The data consisting of all the specified metrics of the 3000 X (Twitter) users that make up the data set were visualized and the data set created from these visuals was added to the data set from Convolutional Neural Networks models; 6 different algorithms have been implemented, such as InceptionRN, MobilenetV2, Imagenet, Xception, Densenet201 and Resnet101. The effect of the structure of the visuals on performance was analyzed through experimental applications on visual types created in different structures.

II. LITERATURE REVIEW

In this part of the article; Articles on the detection and classification of social network users have been examined, the methods adopted in solving the problems focused on in the relevant articles have been examined and their achievements have been presented. Social networks have been the subject of many studies in the literature due to their high number of users as well as their effects on political, social and economic areas. The main issues are the structure of user profiles, interaction styles and the quality of the accounts. Detection of malicious accounts such as spam, bots and identity fraud used in social network manipulation is an important area of work. There are several methods in the literature that have been used to detect malicious social network users, which are discussed in the experimental part of the article. Gannarapu and colleagues presented an automatic mechanism for detecting and removing bots on social media platforms [7]. Their research used bot detection technique based on machine learning algorithms. In the other article, active and passive users, as well as malicious accounts, were classified using machine learning, and a success rate of 96.81% was achieved [8]. In the study by Loyola-gonzález et al., a pattern-based classification

mechanism was used to detect social bots, especially for X (Twitter). A new feature model for bot detection has been introduced. Experimental results have shown that their mechanism outperforms other state-of-the-art classifiers, not based on patterns, and that feature models give better classification results than others reported in the literature [9]. In the vaccine bot detection study conducted by Yuan et al. in the United States, a retweet network about the MMR vaccine after the 2015 California Disneyland measles outbreak was examined and the communication patterns of anti-vaccine and pro-vaccine users and the role of bots on Twitter were investigated. Using supervised machine learning, users were classified into anti-vaccine, vaccine-neutral, and pro-vaccine groups.

It was discovered that pro and anti-vaccine users retweeted mainly from their own opinion groups. Additionally, bot analysis discovered that 1.45% of corpus users were identified as possible bots, accounting for 4.59% of all tweets in the dataset. In the study, it was determined that anti-vaccine social bots caused an environment that hindered the spread of vaccination [10]. In the study by Zago et al., a solution mix was proposed in the form of a proof-of-concept platform that combines the agility of artificial intelligence with the expertise of human analysts to detect and shield against the interference of social bots in detecting social robot activities [11].

In addition to machine learning applications used in bot detection, there are also deep learning applications. In our study, which detects the character of users in social networks and applies deep learning, the classification performance was higher than the machine learning application in the experimental studies conducted on the data set where convolutional neural networks were applied. In this regard, it has been determined that more effective results can be obtained by creating a data set from images with appropriate metrics and applying appropriate parameters and algorithms [12]. In the study by Cai et al., a behavior-enhanced deep model (BeDM) was proposed for bot detection. The proposed model views user content as temporal text data instead of plain text to extract latent temporal patterns. BeDM combines content knowledge and behavioral knowledge using the deep learning method. Experiments conducted on a real-world dataset collected from Twitter have also demonstrated the effectiveness of the proposed model [13].

In the study conducted by Ping et al.; A social bot detection model based on deep learning algorithm (DeBD) has been proposed [14]. The model basically includes three layers. The first layer is the collaborative content feature extraction layer, which focuses on feature extraction of tweet content and the relationship between them. The second layer is the tweet metadata temporal feature extraction layer, which considers tweet metadata as temporal information and uses this temporal information as the input of LSTM to extract the temporal feature of the user's social activity.

The third layer is the feature fusion layer, which combines extracted common content features with ad-hoc features to detect social bots. To evaluate the effectiveness of the DeBD model, experiments were conducted on three different types of new real-world social bot datasets. Experimental results have proven the success of the applied methodology. In the literature review, detailed information about machine learning and deep learning studies used in detecting malicious accounts is given, as well as examining the structure of user profiles in social networks. Since there are no precedent studies that used

a similar method to analyze the effect of the structure of the visuals on performance, it is not given. In this respect, the article has a unique structure.

The difference of this article from the literature and its contribution to the literature can be summarized as follows.

- In the article, an up-to-date and original data set was used in the profile analysis of social network users.
- It is a comprehensive data set because it addresses both the static profile structure and the dynamic interaction of the variables from which the data set is created.
- The originality of the article is the use of a visualization methodology applied on the numerical values in the data set and an alternative and more successful deep learning application method to machine learning.
- The article offers a new approach in terms of measuring the effect of the visualization format on the classification performance with deep learning applied on the data set.

III. EXPERIMENTAL RESULTS

In this section, the process of creating the data set used in the thesis study and the applied methodology are explained. The metrics that make up the data set and the data cleaning, data integration, data reduction, data transformation and visualization processes covered by the applied data mining process are explained. Machine learning and deep learning algorithms used in the applied methodology, learning models, hyperparameters used, analysis and evaluation processes are discussed. Although there are many metrics used in the literature to analyze user profiles and detect bots, the use of an optimal number of but comprehensive metrics is the main criterion. In this sense, the data set of the experimental part of the article study was created from 10 metrics that can comprehensively scale a profile dynamically and statically. The data used was created by ourselves and obtained from current Twitter user data. In this article, those to be tagged are defined in 3 different groups in order to conduct behavioral analysis based on the physical interaction of the focal social media users. These groups are popular-active, observer-passive and spam-bot-malicious categories. These categories were created in accordance with the literature studies and experimental research and were revealed as the most appropriate combination in which social media users can be grouped in general terms.

The following methodology was followed for experimental application in the article:

- a) An original data set consisting of current data and comprehensive variables has been created. The metric data to be used when creating the data set was captured with Twitter Rest API.
- b) By applying data structuring processes such as data cleaning, data integration, data reduction and data transformation on the captured data, the data set is made ready for work.
- c) After the data was filtered, time series were applied on the metric data, the missing parts of the data set were completed and reduced to a single dimension. Min-Max Optimization was applied to eliminate the variance difference between the data.
- d) Numerical values in the data set were converted to visuals. Visualization process was applied to the numerical values of the data set and the data set was brought into a format that can be used with deep learning algorithms.

e) Image classification was performed with Deep Learning algorithms on the data set consisting of images prepared in different formats and experimental results were obtained. Additional techniques have been applied to improve the achieved achievements.

In the article, using 10 different metrics in interaction analysis was deemed sufficient to analyze the dynamic and static structure of a profile. These metrics; tweets, account age, follower rank, average retweet, average likes, diameter, density, reciprocity, centrality, centrality, and modularity.

Descriptions of these metrics and their purposes for use are as follows.

- (a) Tweets: These are messages sent by users to communicate with others.
- (b) Account Age: It refers to the entire period of time an account has been open since its first opening.
- (c) Follower Rank: It refers to the ratio of the number of followed accounts in the accounts to the number of following accounts.
- (d) Average Forwarding: It refers to the situation where a user reshapes another user's message on his/her own profile.
- (e) Average Likes: It is a parameter that expresses how much a social media user likes the messages shared by others.
- (f) Density: It is the ratio of existing connections to the total number of possible connections in a network.
- (g) Centrality: Measures the average degree of centrality of all nodes within a network.
- (h) Reciprocity: Reciprocity is a proportion of ties that exhibit two-way communication (also called reciprocal coupling) in relation to the total number of ties present.
- (i) Diameter: calculates the longest distance between two network participants.
- (j) Modularity: Modularity helps determine whether the clusters found represent different communities in the network.

The data used to create the data set as in Figure 2 was collected from Twitter. The metric data that makes up the dataset was captured with the Twitter Rest API V1.1 search/tweets endpoint using Socialblade and Netlytic platforms.

id	text	user_id	user_name	user_profile_picture	creation_time
15184011234567890	bu seçimler çok önemli...	1234567890	Ali Yılmaz	https://i.pinimg.com/280x280/12/34/56/1234567890.jpg	2023-01-15T10:30:00Z
15184011234567891	Twitter kullanımı artıyor...	1234567891	Fatma Demir	https://i.pinimg.com/280x280/12/34/56/1234567891.jpg	2023-01-15T11:00:00Z
15184011234567892	Yeni teknolojiler hayatımızı...	1234567892	Emre Kaya	https://i.pinimg.com/280x280/12/34/56/1234567892.jpg	2023-01-15T11:30:00Z

Fig. 2 A cross-section of sample user data captured with Twitter REST API.

The data set was created from data collected in a total of 16 periods within a 1-year period. Data from 3000 users were used as training and testing data.

Table 1 A Section from Data Normalized with Min-Max (K: User, M: Metric)

A Section from a Reduced Data Set with Simple Time Series										
Users	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
K1	9192	104	1	1,78	12,82	20	0,002933	0,004575	0,290370	0,606270
K2	11342	84	123	422,095	1208,36	19	0,001229	0,000000	0,403700	0,354600
K3	179136	105	2,57	75	268	19	0,015034	0,227400	0,227400	0,817460
K4	17	129	398,50	33	29	0	0,000000	0,000000	0,000000	0,000000
K5
A Section of Normalized Data with Min-Max Optimization										
K1	0,000243	0,660526	0,0001428	0,000225	0,01055	0,121212	0,002933	0,004575	0,29037	0,60627
K2	0,000299	0,532051	0,1757142	0,053429	0,99046	0,115151	0,001229	0,000000	0,4037	0,35460
K3	0,004736	0,666666	0,0003672	0,009493	0,21967	0,115151	0,002432	0,015034	0,2274	0,81746
K4	0,00045	0,820512	0,0569278	0,004177	0,02377	0,000000	0,000000	0,000000	0,0000	0,00000
K5

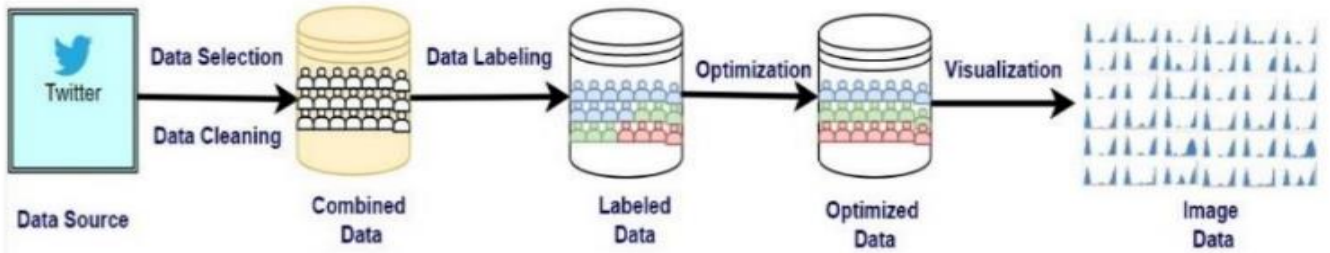


Fig. 3 Methodology used in data visualization for classification with deep learning

The data preprocessing process applied to the data set is given in Figure 3. Preprocessing steps such as cleaning, reduction, transformation and balancing were applied on the data. Since the data set consists of 16 periods of data, time series were applied to obtain images in order to reduce them to a single dimension. The Simple Averages Algorithm of the time series used is given in Equation 1.

$$SMA = \frac{P_m + P_{m-1} + \dots + P_{m-(n-1)}}{n} \quad (1)$$

Min-Max Optimization algorithm was used to eliminate variance differences between metrics in the data set. Thus, the dominance of one metric over other metrics due to variance difference was eliminated. The Min-Max optimization algorithm used is shown in Equation 2.

$$X^* = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2)$$

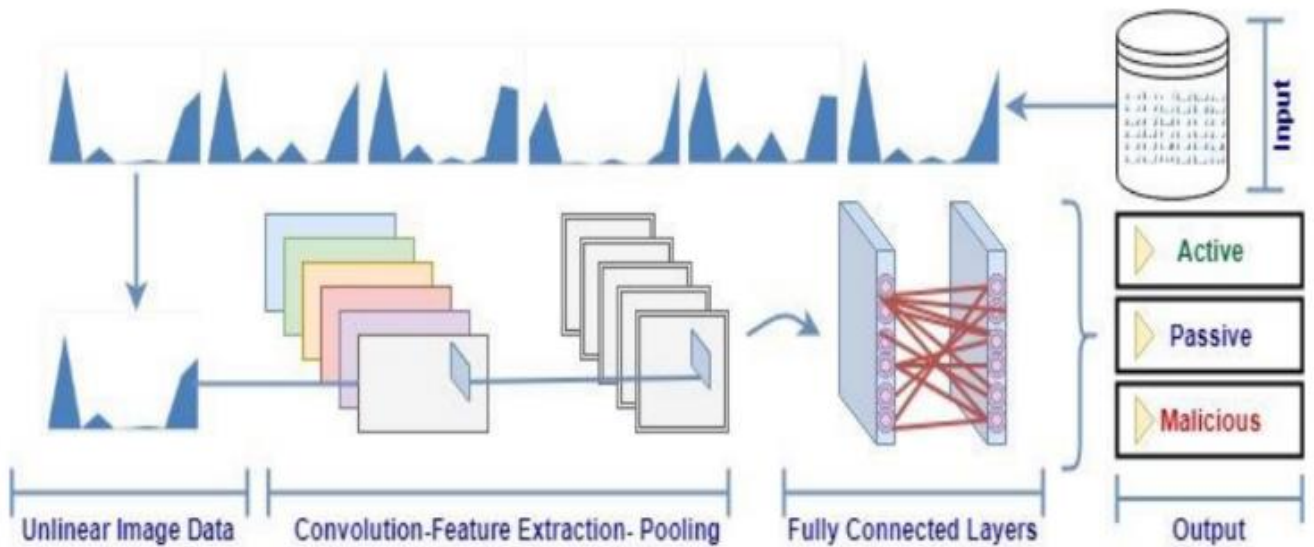


Fig. 4 CNN architecture used in image classification

Table 2 Performance of Deep Learning Algorithms Applied on Images Created in Different Formats

Model	Qrcode	Scatter	Colormap	Histogram	Area Graphics
Xception	71.30	77.21	80.74	87.49	98.67
ResNet101	27.80	31.14	52.91	62.76	98.33
InceptionRN.	23.70	38.60	38.19	73.58	95.58
MobileNetV2	38.40	40.10	47.53	69.74	95.35
ImageNet	29.66	36.29	48.60	74.81	92.67
DenseNet201	34.15	35.43	42.96	65.15	96.67

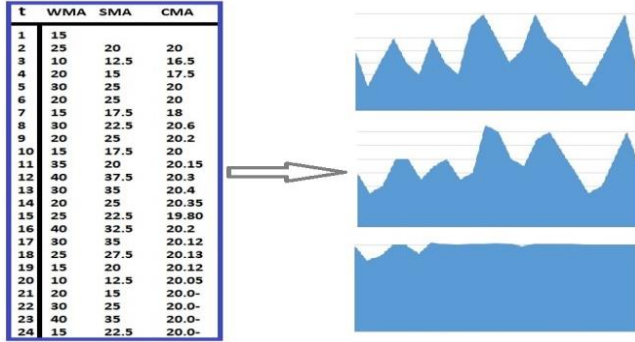


Fig. 5 Visualization of data with time series applied (wma-sma-cma: time series applied, t: time period)

After eliminating the variance difference with Min-Max Optimization, all user data in the data set was visualized by applying the steps shown in Figure 4 in accordance with deep learning and image classification.

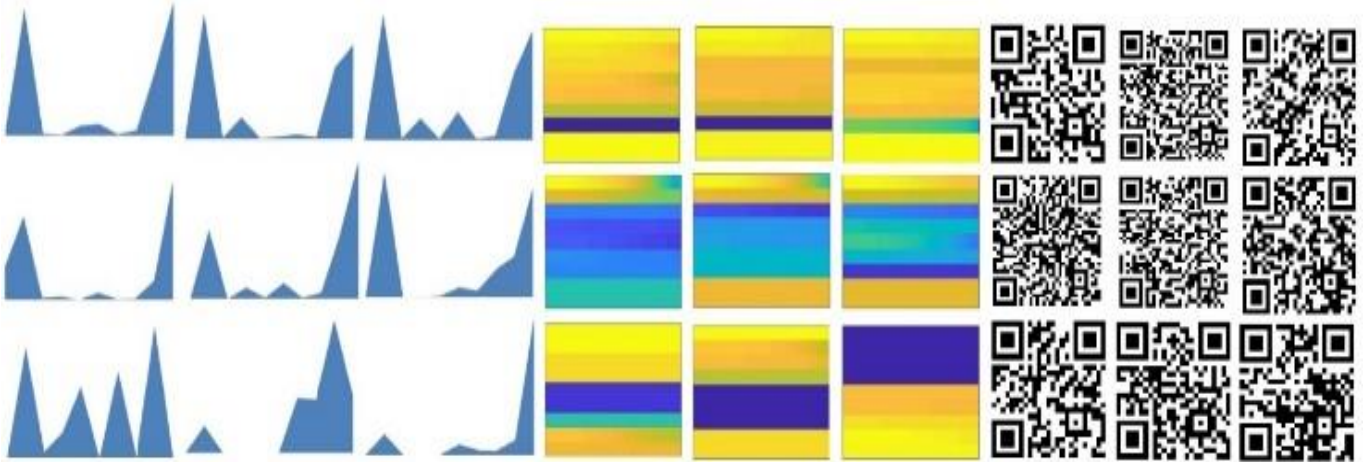


Fig. 6 Examples from the Deep Learning Applied Dataset

These visuals created in Excel look different because they consist of data obtained by applying different time series. Since the images created from the data obtained by simple motion average in the classification methods applied on these images can be better analyzed and classified with high performance rates, the images created using this method have been used in deep learning. Since the images obtained with cumulative motion averages were very similar to each other, they were completely excluded from evaluation.

As shown in Figure 5, the method of obtaining images suitable for the area graphics format applied as a result of the time series used in dimension reduction in order to determine the most suitable format in the visualization process is shown.

In the visualization, images of 3000 users were created in different formats and their classification performance was measured.

The data was reconstructed in visual format as shown in Figure 6 to be processed and classified by the image processing algorithms of deep learning Convolutional Neural Networks.

$$CMA = \frac{X_1 + \dots + X_n}{n} \quad (3)$$

In Equation 3, the formula of the holistic motion average time series used in dimensionality reduction is given.

In visualization, visuals were prepared in formats such as area creation graph, spectrogram graph, colormap, histogram and scatter using Matlab visualization tools. Apart from these, 3000 images compiled from the holistic data set were used to overcome the problems of revealing the performance of the user with qrcode tools due to high resource needs and too much time required in image classification in deep learning.

Experimental applications were carried out on 6 sub-data sets by creating shapes with different structures in all formats for each user data in the data set.

Since hyperparameters such as the number of filters, filter size, activation function, optimization and image pre-processing that make up the convolutional neural networks structure can seriously affect the training success, as many ready-made convolutional neural network models as possible were loaded into the system with the transfer learning module and their performance was tested on the data set.

On the data set with Transfer Learning; Xception, ResNet18, ResNet50, ResNet101, InceptionResNetV2,

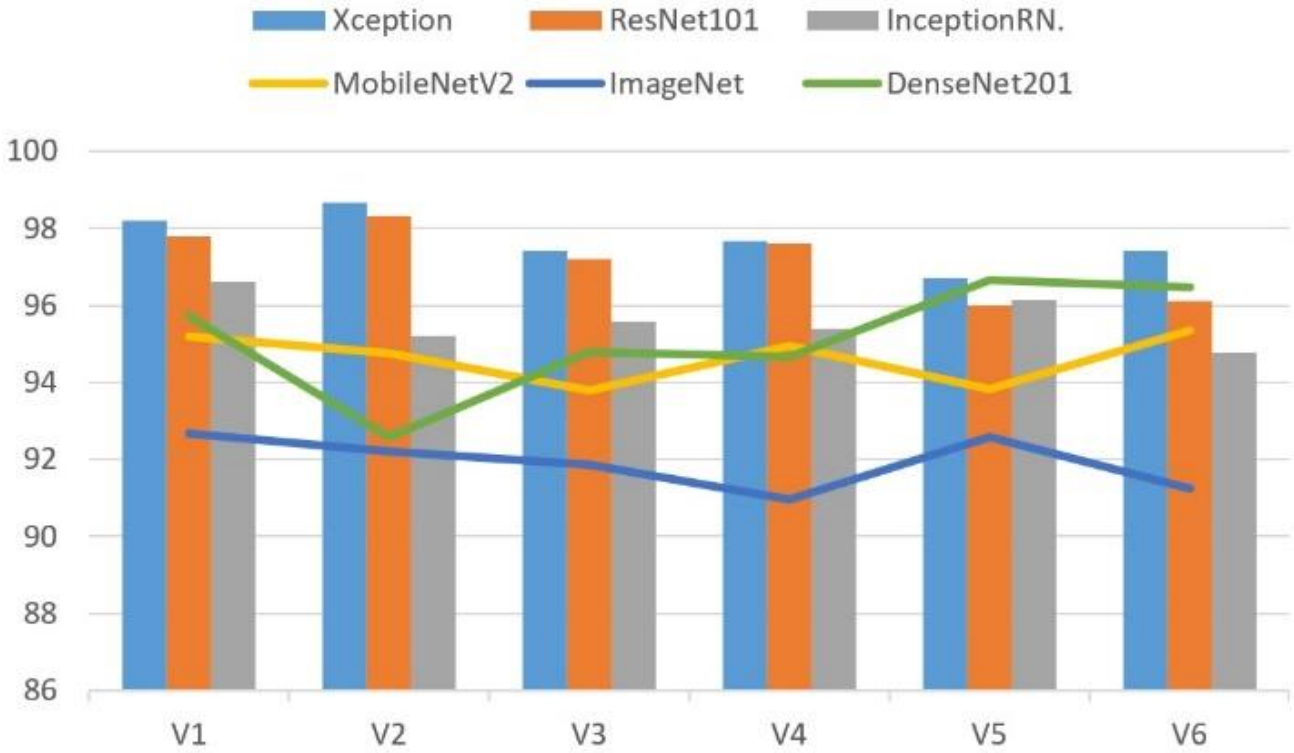


Fig. 7 Graph of Performance Rates in Different Classification Studies Made by Applying Ready-made Convolutional Neural Network Models of Deep Learning (According to Area Graph)

MobileNetV2, ImageNet, DenseNet201, NasnetLarge, InceptionV3, SqueezeNet, AlexNet, VGG16 and VGG19 deep learning architectures have been implemented.

In experimental studies conducted on the data set created from images structured in 5 different formats, results as shown in Table 2 and Figure 7 were obtained. It should be stated that the achievements here are not only due to the visual format, but also that the hyperparameters entered in the transfer learning application used when applying deep learning have a serious impact. But studies have shown that the format of the visuals affects performance at very high rates.

A total of more than 400 training sessions were carried out with these models, which were applied on data sets on multiple platforms with different hyperparameters, and the application results were analyzed.

The classification performance graphs of the Xception algorithm, which showed the highest performance in experimental studies conducted in the Matlab environment with the Dell Precision 7820 Tower workstation, on images of different structures are given below.

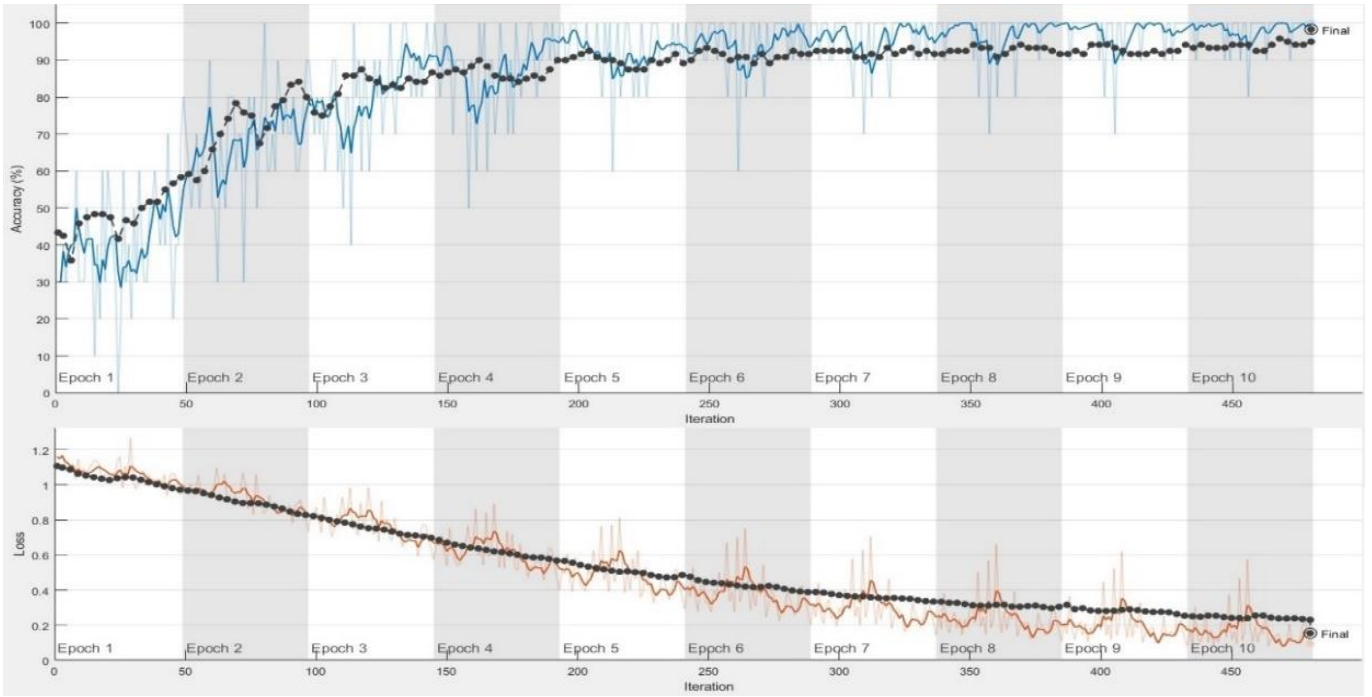


Fig. 8 Accuracy-Loss Graph of Xception Algorithm, Which Performs the Most Successful Classification on Area Graphs

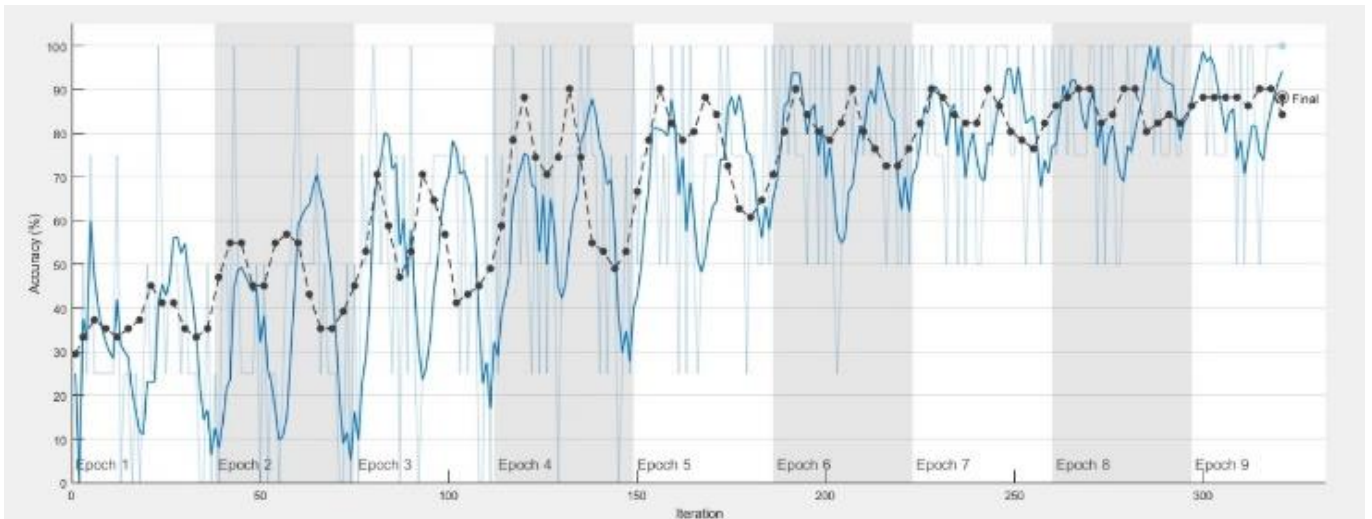


Fig. 9 Data Set Accuracy Graph Created in Histogram Format Applied to Xception

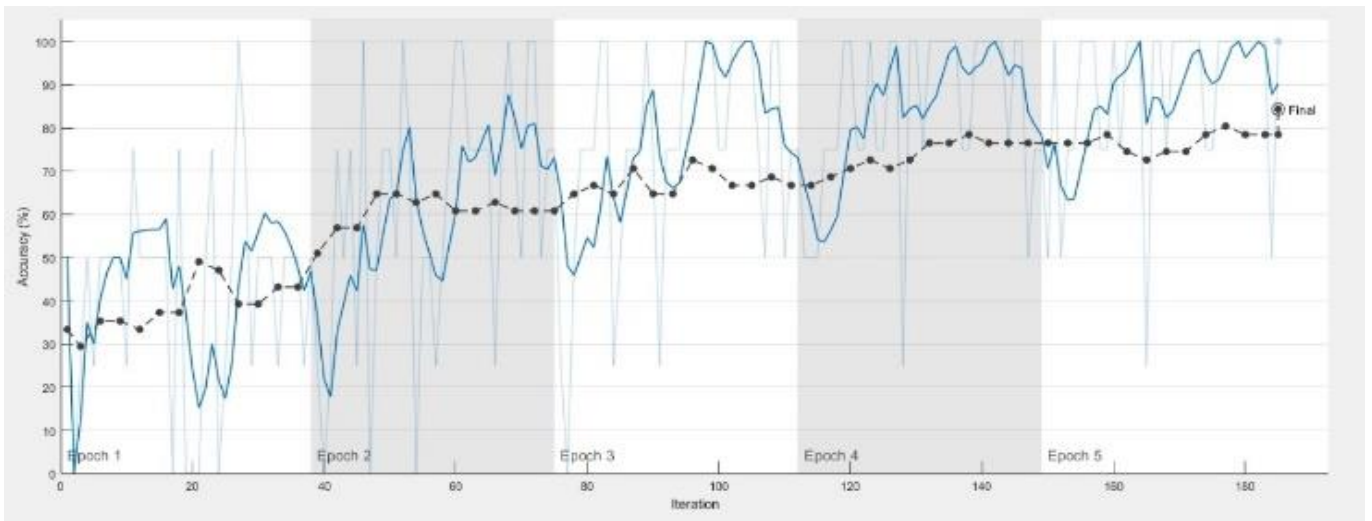


Fig. 10 Data Set Accuracy Graph Created in Colormap Format Applied to Xception

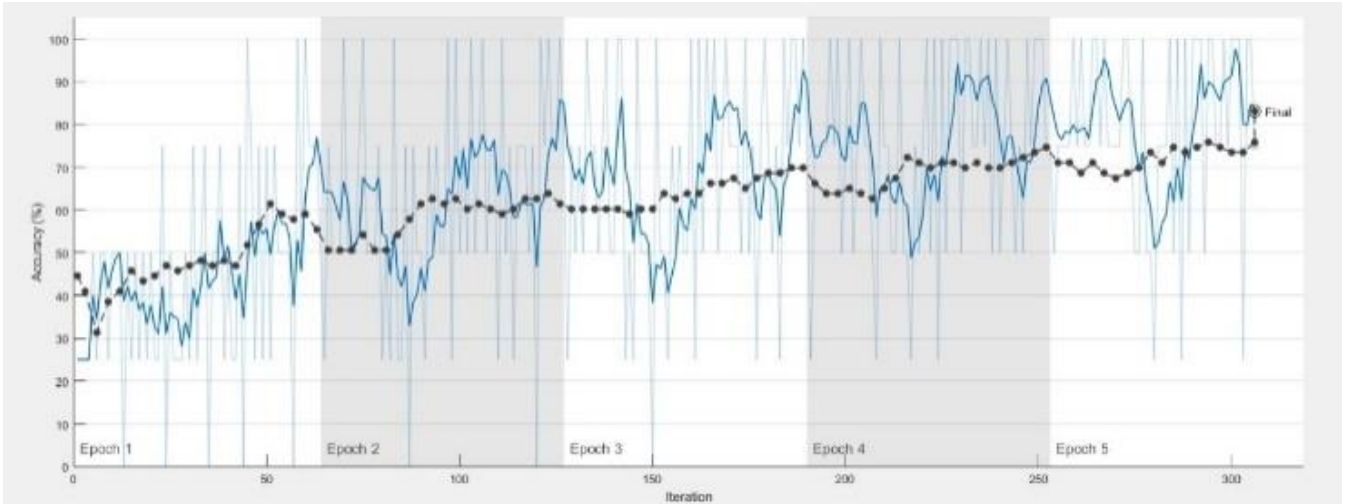


Fig. 11 Data Set Accuracy Graph Created in Scatter Format Applied to Xception

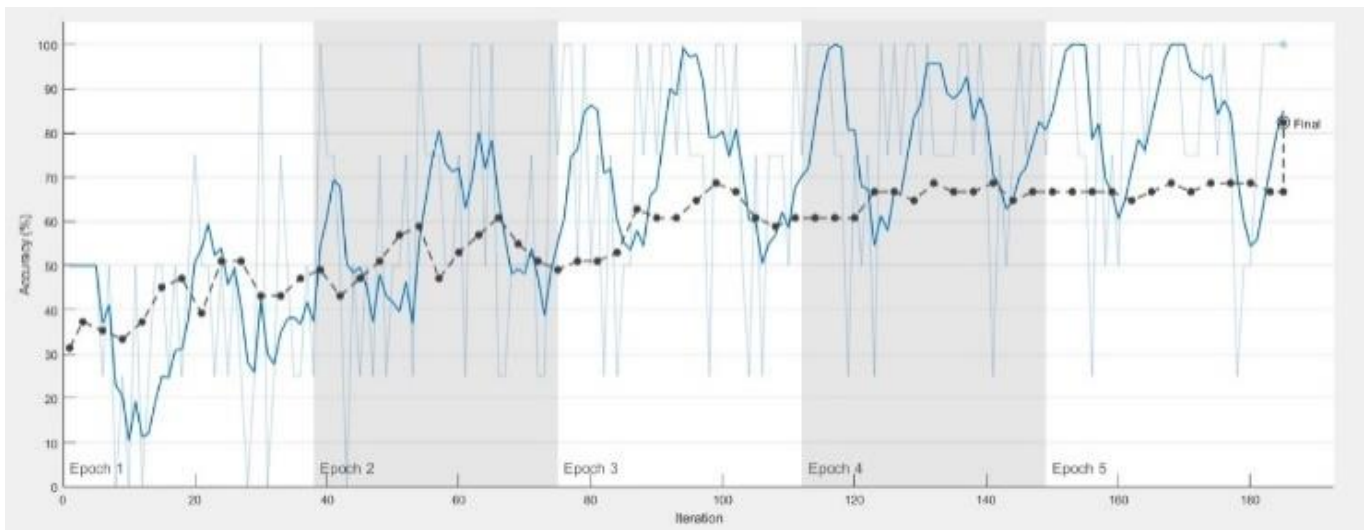


Fig. 12 Data Set Accuracy Graph Created in Applied Qrcode Format

In the experimental study, it was seen that the Xception algorithm gave the most successful classification results for images with all different structures. However, although the Xception algorithm successfully learned and classified area graphs and histogram graphs, it was observed that it interrupted the process after a maximum of 5 epochs because it could not learn from colormap, scatter and qrcode images and could not create a feature map. It has been determined that all other algorithms cannot produce feature maps of images created with different structures other than area and histogram graphics, and therefore cannot make a successful classification.

IV. DISCUSSION

In the article, it was tried to classify social network users as normal or malicious based on their interaction styles. The aim is to detect the source of manipulation in social networks through malicious account detection. Numerical values of the data set were visualized in order to apply deep learning on the data set. Since the desired level of performance could not be achieved in previous interaction analysis studies, this

methodology was developed as an alternative method [15], [16]. It has been observed that the performance in classification with machine learning is lower. As an alternative method, a visualization process was applied and classification was tried to be made with deep learning. In the developed methodology, high performances were achieved with the Xception algorithm used by applying Convolutional Neural Networks. However, as a result of approximately 400 experimental applications, it was determined that the performance depends on the hyperparameters used in the studies carried out with Transfer Learning, and that performing experiments on data sets containing images in a suitable format from which the algorithms can create feature maps affects the performance. It has been determined that in images such as histograms and area graphs, where the area corresponding to each metric is clear, the feature map can be created and thus the source of interaction can be revealed, but in formats where the metric values are melted into the image, the algorithms cannot decipher the structure of the visual and therefore cannot learn the feature map.

V. CONCLUSION

Nowadays, due to their political, social and economic effects, the creation of trust indices of user profiles in social networks; Detecting malicious accounts is valuable in preventing manipulation and disinformation in social networks. In this article, a data set was created using effective and unique metrics that can measure dynamic and static interaction in the social network, and deep learning algorithms were applied on this data set. In the experimental studies carried out, it has been seen through experimental analysis that the creation method and structure of the visuals in the visualization study carried out in order to apply deep learning in the methodological approach applied are of critical importance in terms of deep learning algorithms being able to extract feature maps and make classification. It has been determined that classification studies performed on images that are not in the appropriate format are unsuccessful. However, when algorithms such as Xception were applied on images in appropriate formats, such as images consisting of area graphics suitable for creating feature maps, 98.67% performance could be achieved.

REFERENCES

- [1] Simon Kemp, We Are Social "Digital in 2020" Report, <https://wearesocial.com/digital-2022> (2022).
- [2] Hoaxy-Indiana University, <https://hoaxy.osome.iu.edu/>, (2023).
- [3] H. İş, A. A. Müngen, T. Tuncer and M. Kaya, "Frequent pattern mining for community dedection in web logs group based habit dedection in community using network traces," 2017 International Artificial Intelligence and Data Processing Symposium (IDAP), 2017, pp. 1-7, doi: 10.1109/IDAP.2017.8090293.
- [4] [4]. H. İŞ and T. TUNCER, "Confidence Index Analysis of Twitter Users Timeline," 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), 2018, pp. 1-8, doi: 10.1109/IDAP.2018.8620917.
- [5] [5]. E Ferrara, Manipulation and abuse on social media, arXiv:1503.03752, 2015.
- [6] [6]. R. Rogers, S. Niederer, The Politics of Social Media Manipulation, Amsterdam University Press, 2020.
- [7] [7]. S. Gannarapu, A. Dawoud, R. S. Ali and A. Alwan, "Bot Detection Using Machine Learning Algorithms on Social Media Platforms," 2020 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA), 2020, pp. 1-8, doi: 10.1109/CITISIA50690.2020.9371778.
- [8] [8]. İş, H.; Tuncer, T. Interaction-Based Behavioral Analysis of Twitter Social Network Accounts. Appl. Sci. 2019, 9, 4448. <https://doi.org/10.3390/app9204448>.
- [9] [9]. O. Loyola-Gonzalez, R. Monroy, J. Rodriguez, A. Lopez-Cuevas and J. I. Mata-Sanchez, "Contrast Pattern-Based Classification for Bot Detection on Twitter", IEEE Access, vol. 7, pp. 45800-45817, 2019.
- [10] [10]. X. Yuan, R. J. Schuchard and A. T. Crooks, "Examining Emergent Communities and Social Bots Within the Polarized Online Vaccination Debate in Twitter", Social Media + Society, vol. 5, pp. 205630511986546, 2019.
- [11] [11]. M. Zago, P. Nespoli, D. Papamartzivanos, M. G. Perez, F. G. Marmol, G. Kambourakis, et al., "Screening Out Social Bots Interference: Are There Any Silver Bullets?", IEEE Communications Magazine, vol. 57, pp. 98-104, 2019.
- [12] [12]. İş, H., Tuncer, T. (2021). A profile analysis of user interaction in social media using deep learning. Traitement du Signal, Vol. 38, No. 1, pp. 1-11. <https://doi.org/10.18280/ts.380101>
- [13] [13]. C. Cai, L. Li and D. Zengi, "Behavior enhanced deep bot detection in social media," 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), 2017, pp. 128-130, doi: 10.1109/ISI.2017.8004887.
- [14] [14]. H. Ping and S. Qin, "A Social Bots Detection Model Based on Deep Learning Algorithm," 2018 IEEE 18th International Conference on Communication Technology (ICCT), 2018, pp. 1435-1439, doi: 10.1109/ICCT.2018.8600029.
- [15] [15]. İş, H , Tuncer, T . "Twitter Users' Emotion, Emoticons and Scaling Metrics Based Categorical Interaction Analysis". Journal of Engineering and Technology 2 (2018): 10-18
- [16] [16]. İş, H , Tuncer, T . "Kalite Ölçekleme Kriterleri ile Sosyal Ağ Hesaplarının Etkinliğinin Belirlenmesi". Fırat Üniversitesi Mühendislik Bilimleri Dergisi 31 (2019): 99-108