Journal of Autonomous Intelligence

VOLUME 4
ISSUE 2
2021
ISSN: 2630-5046

Frontier Scientific Publishing

Editorial Board

Editor-in-Chief

Associate Editor

Editorial Board Members

2021 | Volume 4 | Issue 2

Journal of Autonomous Intelligence

http://jai.front-sci.com/

Contents

Journal of Autonomous Intelligence (2021) Volume 4 Issue 2.
doi: 10.32629/jai.v4i2.485

Original Research Article

Spatiotemporal Information Fusion Method of User and Social Media Activity

Chao Yang*, Liusong Yang, Kunlun Qi

School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China

ABSTRACT

Social media check-in data contains a lot of user activity information. Understanding the types of activities and behavior of social media users has important research significance for exploring human mobility and behavior patterns. This paper studies the user activity classification method for Sina Weibo (a very popular Chinese social network service, referred to as "Weibo"), which combines image expression and spatiotemporal data classification technology to realize the identification of the activity behavior represented by the microblog check-in data. Firstly, the user activities represented by the Sina Weibo check-in data are divided into six categories according to POI attribute information: "catering", "life services", "campus", "outdoors", "entertainment" and "travel"; Then, through the Convolutional Neural Network (CNN) and K-Nearest Neighbor (KNN) classification methods, the image scene information and spatiotemporal information in the check-in data are fused to classify the activity behavior of microblog users. The experimental results show that the proposed method can significantly improve the accuracy of microblog user activity type recognition and provide more effective data support for accurately exploring human behavior activities.

Keywords: *Social Media Informatio; Microblog Check-in Data; Classification of User Activities; Machine Learning*

ARTICLE INFO

Received: Jul 21, 2021
Accepted: Sep 1, 2021
Available online: Sep 5, 2021

***CORRESPONDING AUTHOR**

Chao Yang
yangchao@cug.edu.cn;

CITATION

Yang C, Yang L, Qi K. Spatiotemporal information fusion method of user and social media activity. Journal of Autonomous Intelligence 2021; 4(2): 1-11. doi: 10.32629/jai.v4i2.485

1. Introduction

With the development of the Internet and the popularity of social networks, more and more users obtain and share all kinds of social media information on social networks[1]. Social networking software such as Facebook, Twitter and Sina Weibo have more than 3 billion active users worldwide[2], generating and disseminating massive amounts of social media data every day[3]. Sina Weibo is one of the main social media applications in China, by the first half of 2018, the number of Weibo users had reached 337 million, of which 316 million were using smart phones, accounting for 93.5%[4].

Social media data can provide a large amount of information at the level of human and social interaction, especially smart phone users. It can provide social media check-in data with geographical location information, including short text messages, photos, time and location of message release and their corresponding points of interest (POI) information, which can reflect or infer the users' activity content at a certain time and space location to a certain extent[5-6].

Massive social media check-in data can provide data support for tracking and analyzing human movement patterns and behavior patterns in a long time and space scale[7]. Compared with previous location data (such as mobile phone location data), check-in data includes not only time and location information, but also users' behavior information, that

is, information related to social media user activity content; Secondly, the spatiotemporal information sequence of check-in data can be regarded as the spatiotemporal trajectory of users. For example, the check-in point map of 500 million users on foursquare clearly shows the distribution of global human activities[8]. Based on this, more researchers began to explore social media data from different angles. Duan *et al.* proposed a spatiotemporal theme model considering continuous time and regional influence factors, which can more accurately find the evolution of microblog event theme in continuous time[9]; Wang *et al.* proposed the concept of co-word network, extracted the key words of time type to describe different microblog topics, and proved its effectiveness through experiments[10]; Cheng *et al.* explored the activity preference and activity time of Twitter users through the analysis of twitter check-in number in New York City[11]; Rizwan *et al.* used Sina Weibo check-in data to observe the individual check-in behavior and activity intensity in Shanghai[12].

However, social media data is a kind of user generated content (UGC), which is subject to the objective constraints of data acquisition technology, data users' understanding of data and data operation specifications, as well as the subjective influence of location, time and published content. There is uncertainty in data quality[7]. Therefore, when using social media check-in data to analyze the activity category of users, the quality of check-in data needs to be considered[13]. Due to the lack of verification means for the correctness and reliability of social media data, when analyzing user behavior for social media check-in data, we only consider one assumption of data quality: is the POI location of user check-in consistent with the content displayed by users? That is, whether the activity category represented by the POI information is consistent with the multimedia content (text and picture) in the check-in information. As shown in **Figure 1**, we list the cases where the spatiotemporal information and multimedia content contained in Sina Weibo check-in data do not match the activity category represented by the check-in POI information. Therefore, only using the POI information of social media check-in data will directly affect the results of user behavior analysis.

Figure 1. Examples of Sina Weibo check-in data quality problems.

In the research on the behavior classification of microblog users, based on the POI category definition of sina microblog platform, aiming at the main daily activities and contents of social media users, this paper divides the activity types of users into six categories: "catering" "life service", "campus", "utdoor", "entertainment" and "travel". By comprehensively analyzing the spatiotemporal information and image scene information of microblog check-in data[14], the behavior classification of microblog data is realized, so as to improve the accuracy of behavior data classification.

This paper proposes a user activity classification method combining K-Nearest Neighbor (KNN) and Convolutional Neural Networks (CNN). The method determines the temporal and spatial distribution of each check-in data in the three-dimensional space according to the time and space information of the check-in data of Weibo. Based on the theory of KNN model, we believe that the sign in points with similar distribution in time and space are more likely to have the same activity

category; according to the image characteristics of microblog check-in data, the scene information in the image is extracted by using the pre-trained CNN model[15]. We believe that the images uploaded by most users have certain practical expression significance, and the scene information contained in the image can reflect the spatial information or activity state of users. Finally, we use the method of logistic regression to distinguish the type of user activity by combining the probability distribution vector obtained from the above spatiotemporal features and image features. Experiments show that integrating the spatiotemporal information and image scene information of microblog check-in data can significantly improve the accuracy of microblog user activity classification.

2. Microblog check-in data and user activity types

2.1 Microblog check-in data

Every time a microblog user signs in, a microblog check-in data with geographic location information will be generated, including time, location, image, short text and other information; at the same time, when checking in, users will choose to carry the POI information near their location, which not only contains the spatial location information, but also carries the relevant location semantic information, such as the category and name of the POI.

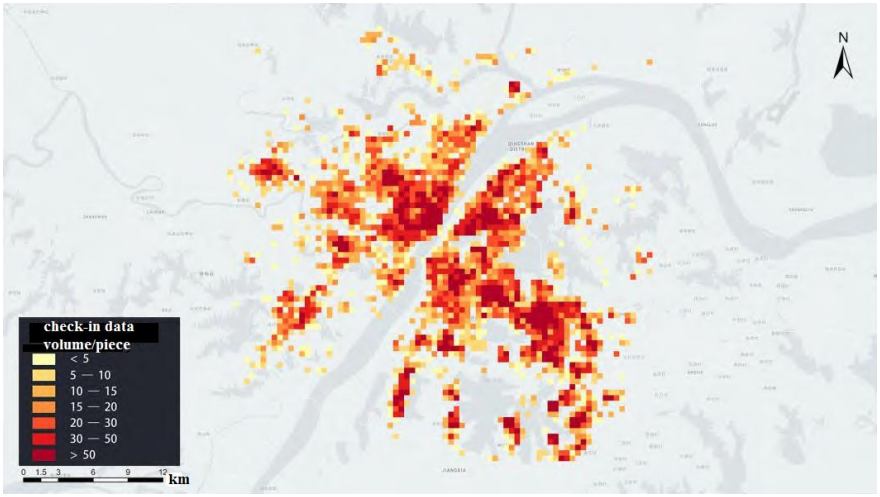

Figure 2. Distribution of Weibo check-in data during January 2015 and June 2015 in Wuhan.

The data set studied in this paper is the check-in data of microblog users in Wuhan from January 2015 to June 2015. As shown in **Figure 2**, the overall data adopts a grid size of 500 m × 500 m to construct a square grid covering the urban area of Wuhan, which is spatially connected with the data of check-in points. The number of check-in points in the grid is used as the classification standard, and the natural breakpoint grouping method is used to select different color bands for each grid according to the sample density; it can be seen that the user's activity area basically covers the main urban areas of Wuhan, while the number of check-in is obviously scarce in the surrounding suburbs and areas with low population activity density.

2.2 Activity types of microblog users

Although the activity type of microblog users' check-in data cannot be obtained directly from the check-in data, most users will choose to carry POI information with high relevance to their activities when signing in. Therefore, it is feasible to classify user activities based on the POI information of check-in data[16]. Sina Weibo officially has 200+ POI categories, such as cafes, cake shops and seafood restaurants, etc. Although this division method is very detailed, for this study, we focus not on the specific activity content of each data, but on the macro activity distribution of user groups. According to the land use regulations and the types of human daily activities, we redefined the original POI types into six categories: "catering", "life service",

"entertainment", "outdoor", "campus" and "travel", among which the POI types corresponding to each activity category and the proportion of microblog data of each category in the check-in data set are shown in **Table 1**.

Table 1. The activity classes of POIss in Weibo check-in data

Activity category	Check-in POI point category	Proportion (%)
Restaurant	Coffee shop, tea shop, dessert shop, fast food restaurant, Chinese restaurant, foreign restaurant, leisure restaurant, pastry shop, cold drink shop, catering food, buffet, barbecue, cooked food, snacks	29.32
Life service	Residential area, community service, dry cleaning shop, photo studio, bank, telecommunications business hall, beauty salon, bath and massage, class III hospital, clinic, drugstore, specialized hospital, 4S store, express, museum, exhibition hall, convention and Exhibition Center	13.38
Entertainment	KTV, billiards hall, bar, game hall, disco, chess and card room, Internet cafe, cinema, theater, nightclub, casino	22.18
Outdoors	Amusement parks, parks, zoos, botanical gardens, city squares, campsites, aquariums, water sports centers, golf related, fitness centers, sports venues, tennis courts, badminton halls, taekwondo halls, natatoriums, basketball venues, football fields, national scenic spots, general scenic spots, scenic spot gates	11.01
Campus	Colleges and universities, adult education, vocational and technical schools, middle schools, primary schools, kindergartens, university centers, school gates, scientific research institutions, libraries	5.01
Travel	Ticket office, airport, railway station, port wharf, long-distance bus station, subway station, hotel guest house, star hotel, Youth Hostel, toll station	19.08

3. Microblog user activity classification model

3.1 Spatiotemporal characteristics based on microblog check-in data

Study the types of user behavior activities, in which time and space factors are very important characteristic factors. Set $S = \{a_1, a_2, \cdots, a_n\}$ as the collection of check-in data samples, where n is the total number of samples in the data set. Each check-in record $a_i (1 \le i \le n)$represents a user's microblog check-in record. For example, after a user completes a microblog check-in behavior, it produces {< 114.35, 30.14 >, 19: 23: 11, *Catering*} such a microblog check-in record, we can understand that the user's time in Beijing is < 19: 23: 11 >, location < 114.35, 30.14 > carried out catering related activities; we denote it as $a_i = \{(l_i, t_i), c_i\} (1 \le i \le n)$, which represents a user's check-in record. Among them, (l_i, t_i)represents the spatiotemporal information of the check-in data, and c_i represents the user activity category information of the check-in data.

Because geographical things or attributes are related in spatial distribution, the closer things in space are more closely related; and in real life, human activities often show a specific time pattern[17]. For example, people often carry out catering activities at a fixed time every day. Therefore, this paper attempts to explain and analyze the user's activities from the perspective of time and space by taking hours as the basic scale and combining with the geographical distribution of spatial features. Based on the above ideas, this paper uses KNN method to model the temporal and spatial characteristics of microblog check-in data, and calculates the classification results of user activities based on the temporal and spatial characteristics of microblog check-in data.

3.2 Image features based on microblog check-in data

In addition to the above-mentioned spatiotemporal characteristics, the microblog check-in data also includes the image features uploaded by users, which can provide feature information related to user activities to a certain extent, supplement and

assist decision-making for information that cannot be directly expressed by spatiotemporal characteristics. For example, for a microblog check-in data whose activity category is unknown and whose space is located around ordinary streets, if the images it carries are mainly images closely related to the catering category, such as dishes, restaurant interior or restaurant door face, we have reason to think that this check-in data probably reflects a user's ongoing catering activities, rather than outdoor category or resident category.

Based on the above ideas, we use the places 365 data set[18] published by MIT in 2015 and its places365 CNN model based on alexnet network training[19]. For each input picture, the model calculates its corresponding scene category probability distribution. Based on the pre-trained places 365 CNN model, combined with manual annotation, this paper maps its original 365 scene categories to the microblog user activity categories described in **Table 1** (as shown in **Figure 3**). For example, we divide the airport and bus station in the original category into travel activities. Cafeteria and bakery are divided into catering activities. Finally, the above model is used to calculate the user activity classification results corresponding to the images in the microblog check-in data (some examples are shown in **Figure 4**).

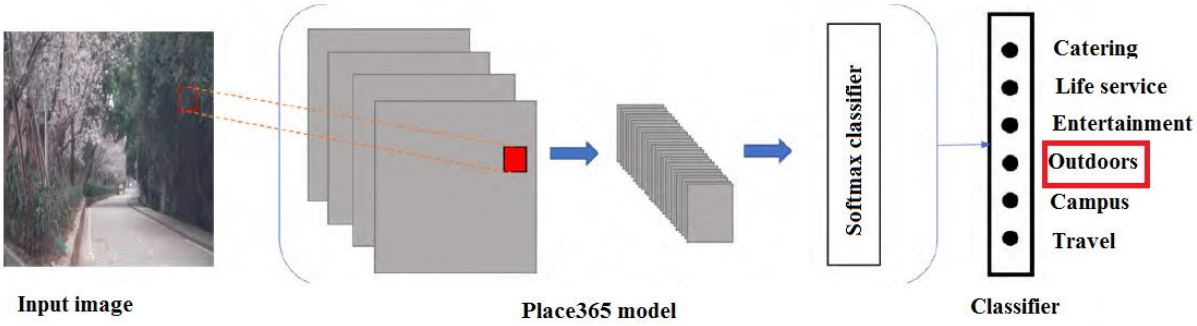

Figure 3. Framework of the image-based user activity classification model.

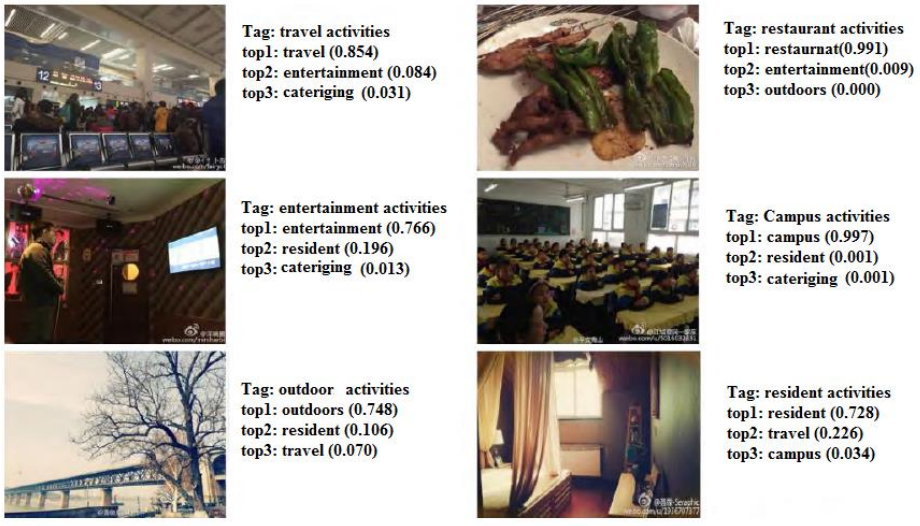

Figure 4. Examples of image-based user activity prediction results.

3.3 Activity classification method based on multi features

For real human activities, most activities will show high aggregation and reciprocation in time and space, such as people's behavior atterns such as three meals a day or working from nine to five, that is, the activity category distribution of a time and space

point can be roughly determined through spatiotemporal characteristics. However, for social media check-in data, there is a lag problem, that is, microblog users may sort out activity photos and publish microblogs after catering or entertainment activities, resulting in the space-time information of check-in data cannot accurately represent the space-time period of activities, resulting in misjudgment of user activity categories based on spatiotemporal characteristics.

We think about two situations: (1) user A has a catering activity in a commercial body at noon, but he only publishes the check-in information in the afternoon. Then, through the obtained spatiotemporal characteristics, we are more likely to think what kind of activity he is doing? (2) User B goes to a KTV in a business at noon to get together with his friends, and clocks in on his social media at the same time. Then, through the obtained spatiotemporal characteristics, we will be more likely to think what activities he is doing? It's not hard to figure out that if we try to judge the user activities behind the social media check-in data only through the spatiotemporal characteristics, we will be more likely to misclassify the above situation (1) into entertainment activities and (2) into catering activities. In view of the above situation, this paper attempts to fuse image features for improvement.

Based on this idea, we propose an activity classification method based on the multidimensional characteristics of social media check-in data:

1) The preprocessed spatiotemporal features are input $\{l_i, t_i, c_i\}$ into the KNN model in the form of triples. For each microblog check-in data, the model based on spatiotemporal characteristics is used to calculate the corresponding probability distribution vector of user activity s_{vector} category;

2) Using the user activity classification model based on image features, the probability distribution vector of user activity category corresponding to the image in each microblog check-in data is p_{vector} calculated;

3) Integrating the probability distribution vector s_{vector} and p_{vector} sum of user activity categories based on spatiotemporal characteristics and image features, a logistic regression equation based on multi-dimensional features of social media check-in data is established: $F = w_1 \times s_{vector} + S_{vector} + w_2 \times p_{vector} + b$;

4) Solve the logistic regression equation to obtain the classification results of user activities based on the multidimensional characteristics of social media check-in data.

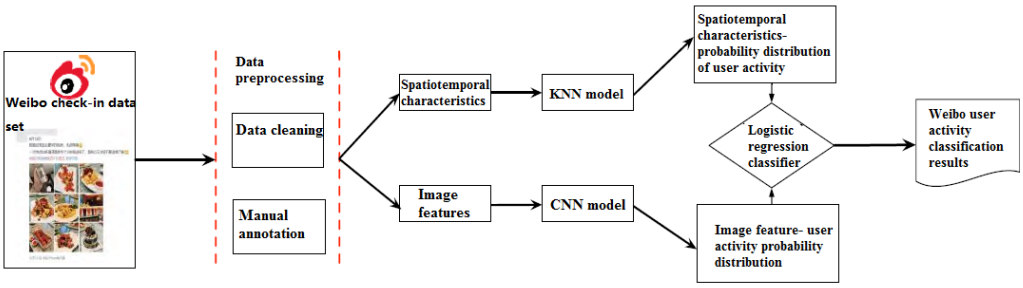

Figure 5. Framework of user activity classification using Weibo check-in data.

4. Experiment and analysis

4.1 Experimental data

The experimental data selected in this paper comes from the microblog check-in data in Wuhan in the first half of 2015, with a total of more than 100,000 pieces of data, which are distributed in all districts of Wuhan. We manually label the data according to the POI, image, timestamp and other information contained in the original data to remove the meaningless or unclear data

Figure 6 shows two microblog check-in data with POI information of different spatial scales. According to the POI information, the microblog data in **Figure 6**(b) can be correctly divided into outdoor activities, but the data in **Figure 6**(a) cannot make a correct judgment. This shows that the POI information contained in the user's sign in microblog data may affect the judgment of their activity category.

Figure 6. The effect on categories of POI at different spatial scale.

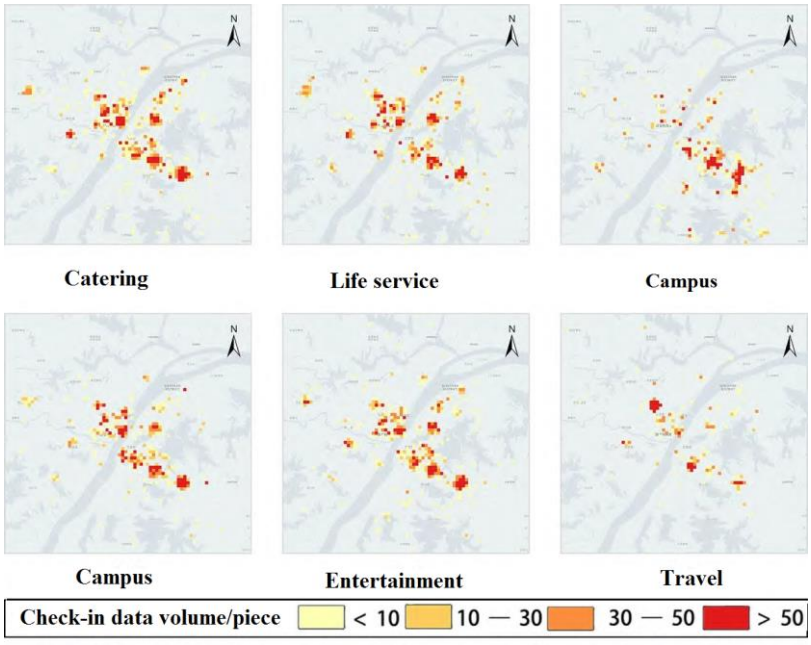

Figure 7. Distribution of user activity classes in Weibo check-in data.

After data preprocessing, there are more than 20,000 effective experimental data used in this paper, of which the data of each activity category accounts for as shown in **Table 1**. **Figure 7** shows the data distribution of different user activity types. It can be seen from the figure that there are a large number of check-in activities in transportation hubs such as Wuhan Tianhe Airport and Wuhan Station, as well as hot areas such as campus and commercial complex, indicating that microblog users prefer to check-in activities when traveling and dining; moreover, the main body of such activities is mainly young people

and students, which is in line with the results of microblog user survey that most of them are highly educated people and young people.

In order to ensure the effectiveness of precision comparison, the data set is randomly divided in this experiment. The proportion of training data and test data is 4:1. The experimental results in this paper are the average value calculated by 10 times of random division.

4.2 Experimental results

According to the above ideas, this paper compares three different methods based on the spatiotemporal characteristics, image characteristics and multi feature fusion of microblog check-in data to verify the effectiveness of the method in the activity classification task of microblog check-in users. As can be seen from **Figure 8**, the overall accuracy of the method based on spatiotemporal characteristics is basically the same as that of the method based on image features; the model based on image features should be relatively robust, and its precision and Kappa coefficient have been significantly improved; the performance of the multi-feature fusion method proposed by us is significantly better than that of the single feature method using only spatiotemporal characteristics or image features. Its accuracy has been improved by nearly 10%, and the recall rate, F1-Score and Kappa coefficient have also been improved by 7%~9% The above results show the effectiveness of this method for the activity classification task of microblog check-in users.

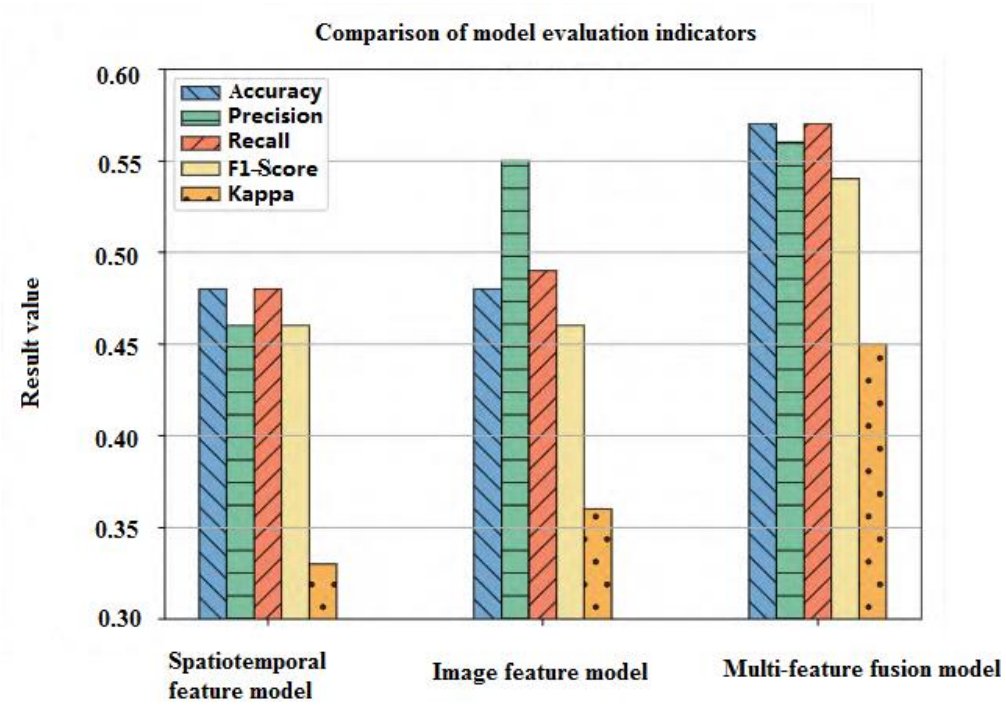

Figure 8. Comparison of three models on different evaluation indicator.

As shown in **Figure 9**, we give the confusion matrices of the three models in each subdivision category. The results show that the three methods show obvious differences in different user activity categories. For the method based on spatiotemporal characteristics, it has nearly 90% accuracy rate in the category of travel activities, which may be because most of the target locations related to travel activities are concentrated in railway stations, airports and other places, which are generally unique in cities. Secondly, the method based on spatiotemporal characteristics also has an accuracy rate of nearly 60% in catering categories. It can be seen from the confusion matrix that a considerable part of the data is misjudged as catering categories. The reason for this phenomenon may be related to the distribution of

urban catering industry. The catering industry is mostly distributed in the interior of major businesses, around major scenic spots or near major universities in order to seek stable passenger flow or more potential customers; however, this also leads to a problem. Assuming that most people choose to release check-in data related to food in a business or around a scenic spot, some data belonging to other categories, such as entertainment or outdoors, will be misjudged as catering activities due to the temporal and spatial aggregation.

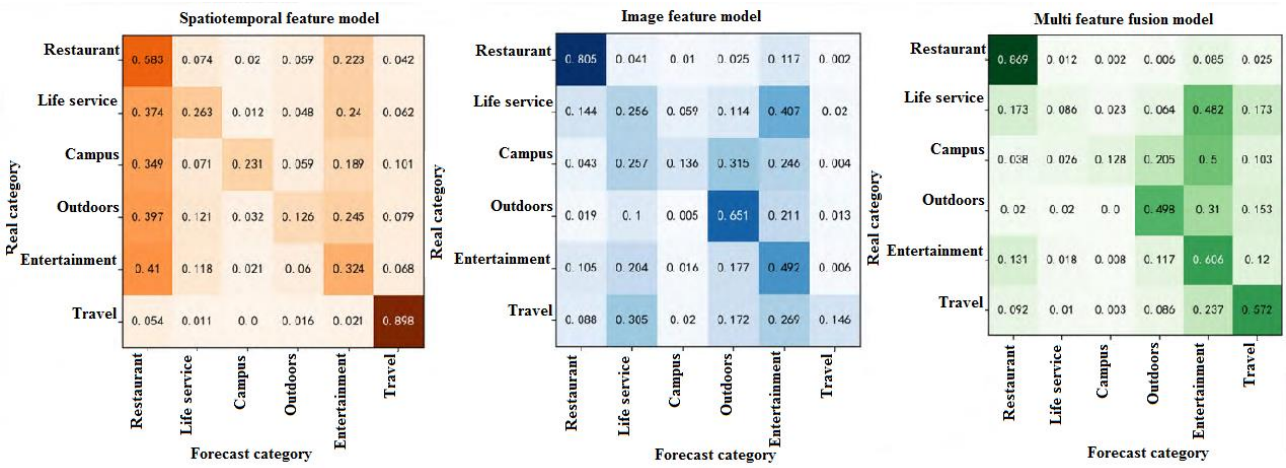

Figure 9. Comparison of the confusion matrix of models.

For the method based on image features, it has shown considerable results in catering category and outdoor activity category, and the accuracy of catering category has reached 80%. This is because most of the images carried by microblog check-in data are related to user activities, such as dishes in catering category, restaurant interior or sky and grassland in outdoor category. The images of these categories have high discrimination and can also provide more information related to user activities; for campus, travel and other categories, the image information carried by these categories, such as the appearance of campus roads and railway stations, does not contain clear feature points and cannot provide strong activity category related information, which also leads to low accuracy of these categories and more misjudgments and misjudgments.

This method can integrate the advantages of different features by integrating space-time and image scene information. While ensuring the classification accuracy of catering category and travel category, it also achieves an overall accuracy of more than 50% in outdoor category and entertainment category. Compared with the single feature method, the accuracy of the multi-feature method in catering and entertainment categories has been improved by 5% and 10% respectively, which shows that this method can alleviate the impact of lag in microblog check-in data to a certain extent; on the other hand, for some microblog check-in data, the single spatiotemporal characteristics image features may not be able to fully express their scene semantics, but the integration of different features can make up for the lack of semantic expression ability of single features in some scenes, so as to improve the accuracy of user activity classification.

By comparing the confusion matrix, we find that although the method based on multi-features has achieved quite good improvement in the overall accuracy compared with the single feature method, its accuracy in some activity categories has decreased to varying degrees. For example, the accuracy of travel categories has decreased by up to 30%. In addition, the method based on multi-features has no significant effect on some activity categories, such as campus category and resident category. This may be because the main influencing factors of these activities are the spatiotemporal characteristics of their activities. For example, the microblog check-in data corresponding to the travel activity category are

mostly located in railway stations or airports, so there is a high correlation in time and space; however, these categories of microblog images are often not the image information corresponding to the activity type, and are often expressed as images of landscape, food or some scenes, such as photos on the subway, at the station or at the airport entrance. These images may be identified as other types, which may interfere with their judgment on the user activity category and affect their classification performance on these user activity categories

Although the classification method based on multi-features cannot comprehensively improve the performance of all user activity types, on the whole, this method can better express microblog user activities, significantly improve the classification accuracy of user activities based on microblog check-in data, and provide more effective data support for accurately exploring human behavior activities.

5. Conclusion

Aiming at the user activity category information hidden in the social media check-in data, this paper uses the machine learning method to decompose the problem, tries to look at and analyze the problem from different angles, and realizes the user activity classification model integrating image and spatiotemporal information. The experimental results show that the method of integrating the spatiotemporal characteristics and image characteristics of microblog check-in data can significantly improve the recognition accuracy of user activity types and provide more effective data support for the research of accurately exploring human behavior activities.

In the future, it will be considered to combine the POI information of microblog check-in data, design a user activity classification model integrating the temporal and spatial characteristics, image characteristics and POI information of check-in data, and realize the end-to-end model from microblog check-in data to user activity category classification, so as to further improve the accuracy of user activity classification method.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Asur S, Huberman BA. Predicting the future with social media. Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01. IEEE Computer Society; 2010. 1: 492–499.
2. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. Business horizons 2010; 53(1): 59–68.
3. Jiang M, Fu KW. Chinese social media and big data: big data, big brother, big profit? Policy & Internet 2018; 10(4): 372–392.
4. Analysis of Chinese Weibo User Data in the First Half of 2018: The Number of Weibo Users in China Reached 337 Million [OL]. Available from: http://www.askci.com/news/chanye/20180821/1 70 6361129537.shtml, 2018
5. Nara A, Tsou MH, Yang JA, *et al.* The opportunities and challenges with social media and big data for research in human dynamics. In: Human Dynamics Research in Smart and Connected Communities. Charm: Springer, Cham; 2018. P. 223–234.
6. Yu Wai T, Yang M, *et al.* Detecting “Hot Spots” of facility POIs based on kernel density estimation and spatial autocorrelation technique. Geomatics and Information Science of Wuhan University 2016; 41(2): 221–227.
7. Tsou MH. Research challenges and opportunities in mapping social media and Big Data. Cartography and Geographic Information Science 2015; 42(sup1): 70–74.
8. The last three months on Foursquare [OL]. Available from: https://foursquare.com/infographics/500million, 2018.
9. Duan L,Guo W, Zhu X, *et al.* Constructing spatio-temporal topic model for microblog topic retrieving. Geomatics and Information Science of Wuhan University 2014; 39(2): 210–213.
10. Wang Y, Fu X, Li M. A new social media topic mining method based on co-word network. Geo-

matics and Information Science of Wuhan University 2018; 43(12): 2287–2294.

11. Cheng Z, Caverlee J, Lee K. You are where you tweet: a content-based approach to geo-locating twitter users. Proceedings of the 19th ACM international conference on Information and knowledge management. ACM; 2010. P. 759–768.
12. Rizwan M, Wan W, Cervantes O, *et al.* Using location-based social media data to observe check-in behavior and gender difference: bringing Weibo data into play. ISPRS International Journal of Geo-Information 2018; 7(5): 196.
13. Crampton JW, Graham M, Poorthuis A, *et al.* Beyond the geotag: situating, 'big data' and leveraging the potential of the geoweb. Cartography and Geographic Information Science 2013; 40(2): 130–139.
14. Lin H, You L, Hu C, *et al.* Prospect of geo-knowledge engineering in the era of spatio-temporal big data. Geomat. Geomatics and Information Science of Wuhan University 2018; 43(12): 2205–2211.
15. Krizhevsky A, Sutskever I, Hinton GE, *et al.* ImageNet classification with deep convolutional neural networks. Neural Information Processing Systems; 2012. p. 1097–1105.
16. Yan L, Duarte F, Wang D, *et al*. Exploring the effect of air pollution on social activity in China using geotagged social media check-in data. Cities 2019;91: 116–125.
17. Huang Q, Wong DWS. Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us? International Journal of Geographical Information Science 2016; 30(9): 1873–1898.
18. Zhou B, Lapedriza A, Xiao J, *et al.* Learning Deep Features for Scene Recognition using Places Database. Advances in Neural Information Processing Systems 27 (NIPS); 2014.
19. Zhou B, Lapedriza A, Khosla A, *et al.* Places: A 10 million image database for scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence; 2018. 40(6):1452–1464.

Journal of Autonomous Intelligence (2021) Volume 4 Issue 2.
doi: 10.32629/jai.v4i2.492

Review Article

The Current Application Status and Expectation of Machine Learning in Unmanned Farm

Baoju Wang[1,2], Yubin Lan[1,2]*, Mengmeng Chen[1,2], Baohu Liu[2,3], GuobinWang[1,2], Haitao Liu[1,2]

[1] School of agricultural engineering and food science, Shandong University of Technology, Zibo 255000, China

[2] National Precision Agriculture International Joint Research Center for aerial pesticide application technology, Shandong University of Technology, Zibo 255000, China

[3] School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo 255000, China

ABSTRACT

With the successful application of machine learning in biological information, face recognition and other fields, it also provides power for the development of unmanned farms. Firstly, this paper expounds the basic concepts of unmanned farm and machine learning. At the same time, it analyzes the application of machine learning in planting and animal husbandry. This paper expounds its application in field weed identification, crop pest detection and crop yield prediction in planting. In animal husbandry, this paper analyzes the application status of machine learning in accurate identification and classification of fish, pigs and other livestock, fish feeding decision-making system and production line prediction of chickens and cattle. It is pointed out that machine learning has some disadvantages, such as difficulties in obtaining and marking training samples, performance defects of embedded chips, and lack of professionals. A general unmanned farm database should be established to study the expert system that can predict the health status of animals and monitor the growth environment of animals in real time. The embedded research of machine learning should be strengthened, and machine learning combined with 5G, big data, sensors and other technologies will become the research direction of unmanned farm in the future. This paper summarizes the application status, problems and prospects of machine learning in unmanned farm, hoping to provide references for further research in the future.

Keywords: *Machine Learning; Unmanned Farm; Crop Management; Animal Husbandry; Accurate Identification; Production Forecast*

ARTICLE INFO

Received: Jul 7, 2021
Accepted: Sep 9, 2021
Available online: Sep 12, 2021

***CORRESPONDING AUTHOR**

Yubin Lan
ylan@sdut.edu.cn;

CITATION

Wang B, Lan Y, Chen M, *et al.* Application status and prospect of machine learning in unmanned farm. Journal of Autonomous Intelligence 2021; 4(2): 12-23. doi: 10.32629/jai.v4i2.492

1. Introduction

Agriculture is one of the most important industries in the world[1], because food is necessary for everyone. It ensures the survival of the global population. The rapid development of agriculture makes the hunger crisis no longer appear[2]. With the trend of growing global population, more agricultural labor force is needed to support agricultural production[3]. However, in the previous decades, the average age of farmers engaged in agricultural work is increasing rapidly. The proportion of people aged 45–64 engaged in agricultural work has increased from 33.48% in 2006 to 43.48% in 2016[4]. Therefore, the current agricultural research mainly focuses on improving higher quality agricultural production with less labor[1].

With the rapid development of information technology, the world's most cutting-edge technologies such as Internet of things (IoT)[5-6], robotics[7-8], big data[9] and artificial intelligence (AI)[10-11] are more and more widely applicated and mature in agriculture, making the concep-

tion of the operation mode of unmanned farm become reality, greatly liberating productivity and improving resource utilization. In the operation mode of unmanned farm, artificial intelligence technology plays the role of thinking and decision-making, and machine learning is one of the most important technologies of artificial intelligence[12].

With the gradual and successful application of machine learning technology in other scientific fields, such as bioinformatics[13], medicine[14], visual tracking[15], robotics[7], climatology[16], remote sensing image processing[17-18], agricultural scientists and scholars pay more and more attention to the application of machine learning in agriculture, which is also the most cutting-edge, modern and promising technology in agriculture[19]. Based on the introduction of the concepts of unmanned farm and machine learning, combined with the practical experience of machine learning technology in the ecological unmanned farm of Shandong University of Technology, this paper summarizes its application status and future development direction in the ecological unmanned farm, so as to provide reference for the better application of machine learning in the unmanned farm in the future.

2. Concepts of Unmanned Farm and Machine Learning Concepts

2.1 Concept of unmanned farm

With the over exploitation of China's agricultural resources, the available cultivated land is decreasing year by year. At the same time, the waste and development without cause or reason of agricultural resources have led to the deterioration of China's agricultural labor environment. Now, China's aging population is becoming more and more serious, there are fewer and fewer labor forces engaged in agricultural labor, and the predicament of no farming is becoming more and more obvious. The in-depth application of information technologies such as Internet of things, cloud computing, big data and artificial intelligence in the agricultural field[20] equipped the unmanned farm with the economic, social and technical conditions.

Unmanned farm is a new agricultural production mode, which does not require too much participation of labor force. Through the joint use of Internet of things, big data, artificial intelligence, the fifth generation (5G) technology and robots and other cutting-edge technologies, all production activities of unmanned farm are carried out with remote control in the whole process, so as to realize the independent operation of equipment, machinery and robots[1].

Unmanned farms use sensor technology to monitor the growth of animals and plants and the working conditions of various production equipment, and use reliable and efficient communication technology to transmit data to the cloud, such as LoRA wireless transmission communication technology[21]. The cloud platform analyzes and processes data through big data technology[22], generates production and operation decisions, then transmits the decision information to the robot, and finally the robot performs specific production activities.

In the unmanned farm, the whole process of agricultural production and operation should achieve accurate management, self decision-making, unmanned operation and personalized service, so as to achieve the sustainable development goal of agricultural production. The architecture of unmanned farm is composed of foundation layer, decision-making layer and application service layer. Its roles and components are described as follows. (1). The foundation layer includes communication system and infrastructure system. (2). The decision-making layer is an intelligent decision-making cloud platform for unmanned farms, which analyzes, processes and stores a large number of data resources and generates decisions. (3). The application layer is the automatic operation equipment system, which uses intelligent agricultural equipment and Internet of things technology. It is the core component of unmanned farm.

The three-tier structure of unmanned farm plays different roles. The basic layer is essential to support the operation of other systems, and the infrastructure system and communication system of the basic layer are responsible for data collection and transmission. The decision-making level implements data man-

agement and makes decisions related to production and operation. The application layer uses machines instead of personnel for production operations. The three-layer structure cooperates with each other to realize the safe and reliable intelligent operation of the unmanned farm[1].

2.2 Machine learning

Machine learning is an important branch of artificial intelligence in the field of computer science. The name of ML was proposed by Samuel[23]. Machine learning is an intelligent method that enables computers to simulate human learning activities, acquire new knowledge, continuously improve performance and realize self-perfection. The basic principle of ML is to construct an algorithm, which can receive data and use statistical technology to predict the output, and update the output when new data is available[24].

Machine learning methods are divided into supervised learning, unsupervised learning and semi-supervised learning. Commonly used algorithms include artificial neural network and deep learning.

2.2.1 Supervised learning

Supervised learning is to obtain an optimal learning model through the training of existing training samples, and then use this learning model to map all inputs into corresponding outputs and make simple judgment on the outputs, so as to achieve the purpose of prediction and classification.

2.2.2 Unsupervised learning

The training samples of unsupervised learning do not have any labeled information. It is to discover the internal relationship of data by learning the training samples without labeled information, so as to provide a basis for further data analysis. It is applicable to scenes that do not have enough previous experience and are not suitable for manual labeling.

2.2.3 Semi-supervised learning

Semi-supervised learning is a collection of supervised learning and unsupervised learning. Part of the data in the training data set is labeled and the other part is unlabeled. A small amount of labeled data and a large amount of unlabeled data are used for learning, so as to obtain the corresponding output. In agriculture, there are usually a large number of unlabeled data due to the limitation of the scene, so the research of semi-supervised learning is very helpful for agriculture.

2.2.4 Artificial neural network

Artificial neural network, abbreviated as neural network (NN), is a mathematical model that simulates the neural system of human brain to process complex information. In fact, it is a complex network composed of a large number of simple components connected with each other. It is a system that can carry out complex logical operation and nonlinear relationship. It is one of the examples of supervised learning[25]. Artificial neuron is the basic information processing unit of artificial neural network operation. The structure of artificial neuron is shown in **Figure 1**.

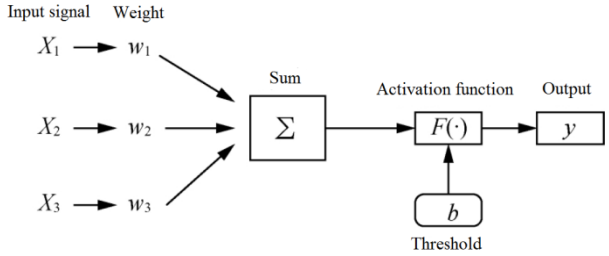

Figure 1. Structure diagram of artificial neural network.

The output of an artificial neuron to an input signal $X = [X_1, X_2 X_3 \ldots X_m]^T$

$$y = f(u + b) \tag{1}$$

$$u = \sum_{i=1}^{m} w_i X_i \tag{2}$$

At present, artificial neural network has been more frequently used in information processing, prediction analysis and other fields[26].

2.2.5 Deep learning

Deep learning is a sub field of machine learning, which is developed on artificial neural network. Its core idea is to automatically extract multi-layer features in the data center through data driven[27] and nonlinear transformation[28]. In essence, it achieves the purpose of feature extraction and transfor-

mation by using nonlinear information processing mechanism and the combination of supervised and unsupervised training, so that the data relationship between samples can be successfully fitted[29]. Deep learning is a deep machine learning model. "Deep" is mainly reflected in the multiple transformation of features[30]. The deep network structure weakens the error features extracted by the previous layer of network to a certain extent, and represents the complex function with fewer parameters, making the network calculation more compact, so as to improve the efficiency and effect[28]. The powerful advantage of deep learning is feature learning, that is, automatically extracting features from the original data and combining lower-level features to form higher-level features[31]. There are many different types of networks for deep learning. The basic networks include deep confidence network, convolutional neural network, recursive neural network, etc.[32], and convolutional neural network is the most widely used in agriculture. The network model leNet-5 is a classical convolutional neural network, and its network structure is shown in **Figure 2**.

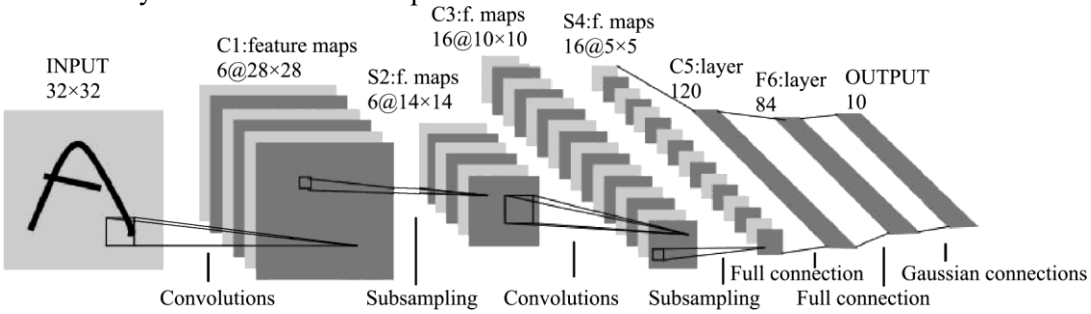

Figure 2. Structure diagram of leNet-5 network.

3. Application of unmanned machine learning on farm

As the key part of unmanned farm is artificial intelligence, and machine learning is one of the key technologies of artificial intelligence, machine learning technology is playing a more and more important role in unmanned farm. This section will discuss the application of machine learning in planting and animal husbandry.

3.1 Application of machine learning in planting industry

3.1.1 Application of machine learning in field weed identification

In agricultural production activities, weeds are inevitable accompanying plants in the field. At present, the main weeding methods used in China are chemical weeding, manual weeding, mechanical weeding, biological weeding, etc. The traditional weeding work is time-consuming and laborious. In today's situation of "no man farming", it is impossible to rely on the traditional weeding technology, so the weeding technology based on machine learning has become more and more important.

Using convolution neural network and deep learning to identify and detect weeds is the most widely used method at present. Andrea *et al.*[33] of Bonn University used convolutional neural network to distinguish corn plants and weeds in the early growth stage of crops, and trained the convolutional neural network with the data set generated in the segmentation stage, and the recognition accuracy reached 97.23%. Jiang Honghua *et al.*[34] improved the convolutional neural network when identifying weeds in the field, adding a binary hash layer behind the full connection layer. By comparing the full connection layer feature code and hash code, they found the labels of K images closest to them and classified them into the category with the highest frequency. This algorithm uses 5000 images for training and 1000 data sets for testing (the proportion of the training set to the test set is 5:1), and the field recognition accuracy is as high as 98.6% and the detection accuracy on other weed data sets was 95.8%. Flores *et al.*[35] from the Department of agriculture and bioengineering of North Dakota State University simulated the field conditions in the greenhouse environment, collected the image shape, color and texture feature values, and used support

vector machine model (SVM), neural network (NN), random forest (RF), GoogLeNet and VGG-16 model for recognition and detection respectively. Finally, the recognition accuracy of VGG-16 model in distinguishing soybean seedlings and corn miscellaneous seedlings reached 96.2%, of which the accuracy is the highest among the above five model methods. Meng Qingkuan *et al.*[36] used deep separable convolution combined with compression and excitation network module to build a lightweight feature extraction unit to replace the VGG-16 network in the standard SSD model and improved the speed of feature extraction. The deep semantic information in the extended network was fused with the shallow detail information. The average detection accuracy of the improved deep learning detection model for corn and weeds is 88.27%. Liu Huili *et al.*[37] built a convolution neural network model with multi-scale hierarchical features based on the deep learning framework tensorflow, and applied the 4 times expanded unit convolution kernel to obtain the recognition model of corn seedling image. Its recognition accuracy reached 99.65%.

In weed management in the field, by improving various machine learning algorithms, the recognition accuracy of weeds has been very high, but most of them are planted and collected in the laboratory, and no field test is carried out in the field. Due to the more complex environment in the field, it will increase the recognition difficulty of machine learning algorithm. We should strengthen the landing experiment and improve the algorithm model through the real field scene, making the machine learning algorithm better applied to the field weed identification project.

3.1.2 Application of machine learning in pest detection

In agriculture, in addition to the great impact of weeds on crops, pest control is another important issue of crop planting. In terms of pest control, the current common practice is to spray chemical agents evenly in the planting area. Although this method is the most effective, the use of chemicals will also cause environmental pollution and threaten environmental safety[38]. Due to the application of in-depth learning in precision agriculture, precision spraying is realized in the process of pest control, and the use of pesticides is reduced.

Pantazi *et al.*[39] used the methods of artistic neural network (ANN) and XY-Fusion network to detect and identify healthy silymarin plants and plants infected by aspergillus niger. Using XY-Fusion method, the detection accuracy reached 95.16%. Ebrahimi *et al.*[40] used SVM classification method to detect thrips on crop canopy images, and used a new image processing technology to detect parasites that may appear on strawberry plants. Support vector machines with different kernel functions were used to classify parasites and detect thrips. The results show that the support vector machine model with regional index and brightness as color index had the best classification effect, and the average error was less than 2.25%. Chung *et al.*[41] proposed a nondestructive method to distinguish infected rice seedlings and healthy seedlings at the age of 3 weeks by using machine vision. They developed a support vector machine (SVM) classifier to distinguish infected and healthy seedlings, and used genetic algorithm to select the basic features and optimal model parameters of SVM classifier. The results show that the proposed method had the accuracy of 87.9%, realized automatic detection of infected plants, improved grain yield, and reduced the time consumption. Zhang Yinsong[42] carried out target detection and recognition of armyworm board pests. He adopted SSD target detection algorithm that can detect in real time, and reduced the problem of small size of pests on the basis of SSD algorithm. Deconvolution was used to realize the feature fusion of high-level and low-level, and then the fused features were used to establish the feature pyramid. Then he detected layer by layer to obtain the optimal recognition model. The results show that the accuracy of model recognition is 91.8%. In terms of the problems of low image recognition accuracy and low model training efficiency of the traditional leNet-5 model in the classification of complex texture images, Liu Zhiyong *et al.*[43] improved the traditional leNet-5. They used the PReLU function as the activation function, added the concept structure module group to the network, adopted the dropout strategy and added batchnormalization, etc., and

proposed the improved leNet-5 model. In the experiment of identifying tomato diseases and pests, its improved model recognition accuracy is as high as 95.3%. Moshou *et al.*[44] identified and detected winter wheat infected with yellow rust, nitrogen stressed plants and healthy plants, and adopted the method based on SOM neural network and hyperspectral reflection imaging. The results show that the accuracy of identifying nitrogen stressed plants was 100%, that of plants infected with yellow rust was 99.92%, and healthy plants was 99.39%.

3.1.3 The role of machine learning in output prediction

Crop yield prediction plays a very important role in unmanned farm operation and is of great significance to improve the production and management level of the farm. You team[45] abandoned the traditional methods used in the field of remote sensing and adopted convolution neural network (CNN) and long-term and short-term memory network (LSTM, a time recursive neural network) to automatically extract relevant features from the original data, and used the deep Gauss process to integrate the spatio-temporal information of the data to evaluate their methods in the task of predicting soybean yield. The results show that their model has a prediction accuracy 15% higher than that of the U.S. Department of Agriculture averagely. Ali *et al.*[46] carried out the work of estimating grassland biomass in small-scale farms with intensive management in Ireland. They adopted the methods of multiple linear regression (MLR), artificial neural network (ANN) and adaptive neuro fuzzy inference system (ANFIS) model, in which the ANFIS model combines the advantages of artificial neural network and fuzzy logic and is evaluated in two intensively managed grassland farms in Ireland. The results show that ANFIS has better effect than the other two methods.

By summarizing the relevant literature of machine learning in the planting industry, it is found that the improved machine learning algorithm has great recognition accuracy and prediction effect, which shows that machine learning can be applied in unmanned farms, but the embedded research of the algorithm should also be strengthened and field tests should be carried out, so that machine learning can be better applied in unmanned farms and promote the intelligent development of unmanned farms faster.

3.2 Application of machine learning in animal husbandry

The main application scenarios of machine learning in animal husbandry are fishing grounds and farms. On the one hand, it is used to accurately identify animals, monitor animal behavior in real time, and provide growth information for producers. On the other hand, machine learning technology is mainly applied to the monitoring of production lines to provide producers with production information to create the greatest economic value.

3.2.1 Application of machine learning in accurate identification of livestock

The intelligent identification of fish by machine learning lays a foundation for further prediction of fishery situation. The accurate prediction data of fishery situation can solve the problem of the lack of fishery standard service based on the standard system in most fishery standard service systems. It can provide data decision-making basis for the revision guide of fishery standards[47], and offer real-time monitoring of fish growth and health data for fishery owners, providing data support for fish farming.

Using deep learning to identify and detect fish is the most common method. Wang Wencheng *et al.*[48] used ResNet50 network to identify and detect turbot, yellowfin snapper, goldfish and mullet, and the test accuracy was more than 96%. They also developed a GUI visual interface using PyQt5. Through the interface operation, the test results were consistent with the prediction category. At the same time, they used DSOD framework to do real-time tracking and detection of underwater targets, and greatly improved the detection accuracy of small targets without losing the detection speed. Yuan Hongchun *et al.*[49] adopted a multi-scale retinal enhancement algorithm (MSRCR) based on Faster R-CNN secondary migration learning and color restoration, which solves the problems of insufficient number of fish samples and rapid detection of blurred fish images. The test results show that the

method useing the network trained by the fish data set with a small number of samples has a detection accuracy of 98.12%. Li Qingzhong *et al.*[50] used the improved YOLO detection algorithm and migration learning to solve the problem of rapid detection of underwater robot fish targets based on video images in a non restrictive environment. Compared with the traditional YOLO algorithm, the improved algorithm improves the detection performance of small targets and overlapping targets, and the detection accuracy reaches 89%. Wang Ye[51] proposed a fish recognition model base on residual network. They used the ResNet50 as basic network, built network by using transfer learning, and introduced attention mechanism, inserting the nonlocal operator of attention mechanism into the residual network in the form of module. The results show that the identification accuracy of the improved network model reaches 98.16%.

With the adoption of intensive management, accurate identification of livestock such as pigs and cattle has become an important issue in farms. Hansen team[52] proposed a non-invasive biometric method for animal face recognition. They tested the method with Fisherfaces, VGG-Face pre- trained face convolution neural network (CNN) model and their own CNN model, and they trained it with artificially expanded data sets. The results show that the recognition accuracy of their own CNN model reached 96.7%.

Accurate identification and classification of livestock plays an important role in animal husbandry. In the research of livestock identification in recent years, scholars have improved the machine learning algorithm, which has reached a very high recognition accuracy and laid a solid foundation for livestock behavior identification and health monitoring.

3.2.2 Application of machine learning in livestock production prediction

Machine learning has the ability to detect and warn problems early, which plays a very important role in animal husbandry. It can monitor poultry in real time, find problems in the production process in time, and take timely actions to avoid these problems and reduce economic losses. Real time production monitoring of animals in farms can timely adjust production strategies and maximize benefits. At present, machine learning is widely used in this field. Morales team[53] used the egg production data of 478,919 hens on the farm and the method of support vector machine to find the problems in the egg production curve. This technology can send an alarm one day in advance to warn that there are problems in the production curve, and the accuracy rate is 98.54%. Alonso *et al.*[54] used the support vector machine regression method to predict the weight of beef cattle a few days before slaughter, and measured 144 animals for 390 times. The average absolute error was 4.27% of the real value.

3.2.3 Application of machine learning in livestock feeding decision

In aquaculture, fish feeding is of great significance not only to reduce the cost, but also to improve the yield of fish. Zhou *et al.*[55] used the near infrared computer vision and neuro fuzzy feed control method to realize the purpose of automatic feeding according to the appetite of fish. The test results show that the accuracy of feeding decision of the model reaches 98%. Zhao Jian[56] monitored the local sudden behavior of fish herds that can characterize the hunger degree of fish herds in circulating water culture. They adopted recursive neural network, particle advection scheme and improved motion influence map. The experimental results show that the average detection accuracy reached 98.91% and the average recognition accuracy reached 89.89%.

The application of machine learning in animal husbandry has shown good results. It has a very good performance in accurate classification and identification, production prediction and feeding decision-making. Due to the need for more rigorous livestock breeding to ensure the accuracy of livestock information, the current machine learning can not completely solve the problems in artificial breeding, but the performance of machine learning can provide better information decision-making support for the breeding process. In the future technological development, machine learning will have a broader application field in the aq-

uaculture industry.

4. Discussion and Prospect

After summarizing the above literature, machine learning technology mainly focuses on machine vision, using machine learning algorithm to detect target objects, but there are also some disadvantages in these applications.

1). The use of machine vision requires a large number of data sets for model training and verification. In the current environment of agricultural machine learning, there is no universal data set, and each experimental team collects the labeled data set by itself. Due to the influence of farmland environment, collecting the data set is time-consuming and labor-consuming, and the labeled data set needs manual labeling by professionals, so that there is no data set with a large amount of circulating data. This not only limits the detection accuracy of machine vision model, but also increases the application difficulty of machine learning in agriculture.

2). At present, machine learning needs high-quality hardware to meet its computing ability, but the current embedded chip has problems such as insufficient computing ability and slow computing speed. The performance breakthrough of embedded chip involves problems in other fields, and it is difficult to have greater research breakthrough in the short term.

3). The application of machine learning in agriculture requires both agricultural production experience and professional knowledge of machine learning. However, there is a serious shortage of professionals with both, which limits the development of machine learning technology in agriculture.

In the operation mode of unmanned farm, machine learning is essential. In view of the application status of machine learning in unmanned farm, we should also strengthen the research on the following aspects.

1). Nowadays, machine learning technology is mainly used in crop field weed management and pest detection. It has been widely used in crops, saving a lot of human, material and financial resources, but it is still less used in fisheries, cattle farms and pig farms. In the future, we should do more researches on animals. Using machine learning technology to dynamically monitor the growth status of animals, and using big data technology, combined with the production experience of experts, we can predict the health status of animals through their daily behavior, and establish a set of expert system, which can timely avoid the large-scale spread of animal diseases. At the same time, machine learning should also be used to monitor the growth environment of animals in real time, so as to provide decision support for improving the growth environment of animals.

2). At the same time, a set of high-quality database should be established in the unmanned farm. The database plays a very important role in machine learning. Therefore, a high-quality database will accelerate the application process of machine learning and greatly improve the management efficiency of the unmanned farm, which is of great positive significance to the construction of the unmanned farm.

3). With the rise of unmanned farm, it has produced a large amount of data in the field of production and intelligent equipment. Machine learning combined with 5G, sensors, big data and other technologies are used to transmit, integrate, process and apply these data, so as to build the farm management system into a real AI system.

4). In view of the problem that most of the current machine learning research are limited to the laboratory, we should strengthen the embedded research of machine learning technology and truly conduct the research in the laboratory in the field. We should strengthen the research on machine learning algorithm, reduce its dependence on the performance of embedded chip, and speed up the training speed and running speed of the algorithm, so as to speed up the application of machine learning in unmanned farms.

In short, the application of machine learning in unmanned farms will have a broader world.

5. Conclusion

This paper summarizes and introduces the relevant literature on the application of machine learning in agriculture in recent years. withdrawing

on the practical experience of ecological unmanned farm in Shandong University of Technology, this paper expounds the application of machine learning in field weed identification, crop pest detection and crop yield prediction in planting industry, and its application in the accurate identification and classification of fish, pigs, sheep and other livestock in animal husbandry, fish feeding decision-making system and production line prediction of chickens and cattle. By summarizing the above literature and practical experience, it is concluded that machine learning has disadvantages in the application of unmanned farm, and there are considerable problems in data sets, professionals and embedded systems. Next by summarizing their own practical experience and current research level and problems in unmanned farms, this paper puts forward the development trend of machine learning in unmanned farms, mainly including the establishment of efficient databases, the establishment of "expert systems", and embedded research combined with multi-domain technologies and algorithms.

The application of machine learning in unmanned farms is developing rapidly, and it is also valued by more researchers. More machine learning technologies will be applied to unmanned farms to realize real unmanned operation and promote the rapid and sustainable development of agriculture.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Wang T, Xu X, Wang C, *et al.* From smartfarming towards unmanned farms: A new mode of agricultural production. Agriculture 2021; 11(2): 145. doi: 10.3390/agriculture11020145.
2. Gollin D, Parente S, Rogerson R, *et al.* The role of agriculture in development. American Economic Review 2002; 92(2): 160–164.
3. Hunter MC, Smith RG, Schipanski ME, *et al.* Agriculturein 2050: Recalibrating targets for sustainable intensification. Bioscience 2017; 67(4): 386–391. doi: 10.1093/biosci/bix010.
4. Yang W. Research on the impact of agricultural labor ageing on agricultural labor productivity (in Chinese). Beijing: Beijing Jiaotong University; 2019.
5. Akbar MO, Khan M, Ali MJ, *et al.* IoT for development of smart dairy farming. Journal of Food Quality 2020; (2): 1–8. doi: 10.1155/2020/4242805.
6. Ramli MR, Daely PT, Kim DS, *et al.* IoT-based adaptive network mechanism for reliable smart farm system. Computers and Electronics in Agriculture 2020; 170: 105–287. doi: 10.1016/j.compag.2020.105287.
7. Takahashi K, Kim K, Ogata T, *et al.* Tool-body assimilation model considering grasping motion through deeplearning. Robotics and Autonomous Systems 2017; 91: 115–127. doi: 10.1016/j.robot.2017.01.002.
8. Gastaldo P, Pinna L, Seminara L, *et al.* A tensor-based approach to touch modality classification byusing machine learning. Robotics and Autonomous Systems 2015; 63: 268–278. doi: 10.1016/j.robot.2014.09.022.
9. Kamilaris A, Kartakoullis A, Prenafeta-Boldú FX. A review on the practice of big data analysis in agriculture. Computers and Electronics in Agriculture 2017; 143: 23–37. doi: 10.1016/j.compag.2017.09.037.
10. Chlingaryan A, Sukkarieh S, Whelan B. Machine learning approaches for crop yield prediction and nitrogenstatus estimation in precision agriculture: A review. Computers and Electronics in Agriculture 2018; 151: 61–69. doi: 10.1016/j.compag.2018.05.012.
11. Konstantinos L, Patrizia B, Dimitrios M, *et al.* Machine learningin agriculture: A review. Sensors 2018, 18(8): 26–74. doi: 10.3390/s18082674.
12. Zhang Q, Yang LT, Chen Z, *et al.* A survey on deep learning for bigdata. Information Fusion 2018; 42: 146–157. doi: 10.1016/j.inffus.2017.10.006.
13. Kong L，Zhang Y，Ye ZQ，*et al.* CPC： Assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Research 2007; 35(Web Server issue): W345–9. doi: doi: 10.1093/nar/gkm391.
14. Kang J, Schwartz R, Flickinger J, *et al.* Machine learning approaches for predicting radiation therapy outcomes: A clinician's perspective. International

Journal of Radiation Oncology Biology Physics 2015; 93(5): 1127–1135.

15. Zhu Y, Zhao J, Wang Y, *et al.* A review of human action recognition based on deep learning (in Chinese). Acta automatica Sinica 2016; 42(6): 848–857. doi: 10.16383/j.aas.2016.c150710.
16. Cramer S， Kampouridis M，Freitas A，*et al.* An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. Expert Systems with Applications 2017; 85: 169–1811. doi: 10.1016/j.eswa.2017.05.029.
17. Huang H, Deng J, Lan Y, *et al.* A fully convolutional network for weed mapping of unmanned aerial vehicle (UAV) imagery. PloSOne 2018; 13(4): e0196302. doi: 10.1371/journal.pone.0196302.
18. Weiss M, Jacob F, Duveiller G. Remote sensing for agricultural applications: A meta-review. Remote Sensing of Environment 2020; 236. doi: 10.1016/j.rse.2019.111402..
19. Kamilaris, Andreas, Pren af eta-Boldu, et al. Deep learning in agriculture: A survey. Computers and Electronics in Agriculture 2018; 147: 70–90. doi: 10.1016/j.compag.2018.02.016.
20. Zhang Y, Niu M, Liu L, et al. A preliminary study on the emergence and development of unmanned farms in China (in Chinese). Agricultural Engineering Technology 2020; 40(21): 27–28. doi: 10.16815/j.cnki.11-5436/s.2020.21.003.
21. Wang H, Hao W, Xia Y, et al. Design of automatic cultivation system for greenhouse crops based on LoRa technology (in Chinese). Microcontrollers & Embedded Systems 2021; 21(2): 71–74, 78.
22. Wolfert S, Ge L, Verdouw C, et al. Big data in smart farming—A review. Agricultural Systems 2017; 153: 69–80. doi: 10.1016/j.agsy.2017.01.023.
23. Samuel AL. Some studies in machine learning using the game of checkers. IBM Journal of Research and Development 1959; 3(3): 210–229. doi: 10.1147/rd.33.0210.
24. Ouf N S. A review on the relevant applications of machine learning in agriculture. IJIREEICE 2018; 6(8): 1–17.
25. Mishra S, Mishra D, Santra GH, et al. Applications of machine learning techniques in agricultural crop production: A review paper. Indian Journal of Science and Technology 2016; 9(38): 1–14. doi: 10.17485/ijst/2016/v9i38/95032.
26. Mou W. Application of machine learning technology in modern agriculture (in Chinese). Electronic Technology and Software Engineering 2018(18): 240–241.
27. Guo X, Tai H. The application and prospect of deep learning in field planting (in Chinese). Journal of China Agricultural University 2019; 24(1): 119–129.
28. Yu B, Li S, Xu S, *et al.* Deep learning: The key to open the age of big data. Journal of Engineering Studies 2014; 6(3): 233–243.
29. Bengio Y. Learning deep architectures for AI. Foundations and Trends in Machine Learning 2009; 2(1): 1–127. doi: 10.1561/2200000006.
30. Fu L, Song Z, Wang D, *et al.* Research progress and application status of deep learning methods in agricultural information (in Chinse). Journal of China Agricultural University 2020; 25(2): 105–120.
31. LecunY, Bengio Y, Hinton G. Deep learning. Nature 2015; 521(7553): 436–444. doi: 10.1038/nature14539.
32. Duan Y, Lv Y, Zhang J, *et al.* Research status and prospect of deep learning in the field of control. Acta Automatica Sinica 2016; 42(5): 643–654. doi: 10.16383/j.aas.2016.c160019.
33. Andrea CC, Daniel BBM, Misael J. Precise weed and maize classification through convolutional neuronal networks. 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM); 2017; Salinas. IEEE; 2017. p. 1–6.
34. Jiang H, Wang P, Zhang Z, *et al.* Rapid identification of weeds in corn fields based on convolutional network and hash code (in Chinse). Transactions of the Chinese Society for Agricultural Machinery 2018; 49(11): 30–38.
35. Paulo F, Zhang Z, Jithin M, *et al.* Distinguishing volunteer corn from soybean at seedling stage using images and machine learning. Smart Agriculture, 2020; 2(3): 61–74. doi: 10.12133/j.smartag.2020.2.3.202007-SA002.
36. Meng Q, Zhang M, Yang X, *et al.* Recognition of corn seedlings and weeds based on lightweight convolution combined with feature information fusion (in Chinese). Transactions of the Chinese Society for Agricultural Machinery 2020; 51(12):

238–245, 303.

37. Liu H, Jia H, Wang G, *et al.* Recognition method and experiment of corn seedling stalk based on deep learning and image processing (in Chinese). Transactions of the Chinese Society for Agricultural Machinery 2020; 51(4): 207–215.
38. Ying R, Zhu Y. The impact of agricultural technology training methods on farmers' agricultural chemical input use behavior: Evidence from experimental economics (in Chinese). China Rural Survey 2015(1): 50–58, 83, 95.
39. Pantati XE,Tamouridou AA, Alexandridis TK, *et al.* Detection of silybum marianum infection with microbotryum silybum using VNIR field spectroscopy. Computers and Electronics in Agriculture 2017; 137: 130–137. doi: 10.1016/j.compag.2017.03.017.
40. Ebrahimt MA,Khoshtaghaza MH, Minaei S, *et al.* Vision-based pest detection based on SVM classification method. Computers and Electronics in Agriculture 2017; 137: 52–58. doi: 10.1016/j.compag.2017.03.016.
41. Chung CL, Huang KJ, Chen SY, *et al.* Detecting Bakanae disease in rice seedlings by machine vision. Computers and Electronics in Agriculture 2016; 121: 404–411. doi: 10.1016/j.compag.2016.01.008.
42. Zhang Y. Identification and counting of pests in sticky board images based on deep learning (in Chinese). Xuzhou: China University of Mining and Technology; 2019.
43. Liu Z, Zhang L, Zhong T, *et al.* Research on identification of tomato pests and diseases based on improved leNet-5 (in Chinese). Journal of Gannan Normal University 2020; 41(6): 70–74. doi: 10.13698/j.cnki.cn36-1346/c.2020.06.015.
44. Moshou D, Bravo C, Wahlen S, *et al.* Simultaneous identification of plant stresses and diseases in arable crops using proximal optical sensing and self-organising maps. Precision Agriculture 2006; 7: 149–164. doi: 10.1007/s11119-006-9002-0.
45. You J, Li X, Low M, *et al.* Deep gaussian process for crop yield prediction based on remote sensing data. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence; 2017 Feb; San Francisco. AAAI Press; 2017. p. 4559–4565.
46. Ali I, Awkwell F, Dwyer E, *et al.* Modeling managed grassland biomass estimation by using multitemporal remote sensing data—A machine learning approach. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2017; 10(7): 3254–3264. doi: doi: 10.1109/JSTARS.2016.2561618.
47. Yu H, Feng Y, Li H, *et al.* Establishment of a systematic service and assistant decision-making system for fishery standard (in Chinese). Journal of Dalian Ocean University 2019; 34(2): 260–266. doi: 10.16535/j.cnki.dlhyxb.2019.02.016.
48. Wang W, Jiang H, Qiao Q, *et al.* research on the algorithm of fish recognition and detection based on deep learning (in Chinese). Information Technology and Network Security 2020; 39(8): 57–61, 66. doi: 10.19358/j.issn.2096-5133.2020.08.011.
49. Yuan H, Zhang S. An underwater fish target detection method based on Faster R-CNN and image enhancement (in Chinese). Journal of Dalian Ocean University 2020; 35(4): 612–619. doi: 10.16535/j.cnki.dlhyxb.2019-146.
50. Li Q, Li Y, Niu J. Real-time detection of underwater fish targets based on improved YOLO and transfer learning (in Chinese). Pattern Recognition and Artificial Intelligence 2019; 32(3): 193–203. doi: 10.16451/j.cnki.issn1003-6059.201903001.
51. Wang Ye. Research on fish recognition based on deep learning (in Chinese). Shanghai: Shanghai Ocean University; 2020.
52. Hansen MF, Smith ML, Smith LN, *et al.* Towards on-farm pig face recognition using convolutional neural networks. Computers in Industry 2018; 98: 145–152. doi: 10.1016/j.compind.2018.02.016.
53. Morales IR, Cebrian DR, Blanco EF, *et al.* Early warning in egg production curves from commercial hens: A SVM approach. Computers and Electronics in Agriculture 2016; 121: 169–179. doi: 10.1016/j.compag.2015.12.009.
54. Alonso J, Castanon AR, Bahamonde A. Support vector regression to predict carcass weight in beef cattle in advance of the slaughter. Computers and Electronics in Agriculture 2013; 91: 116–120. doi: 10.1016/J.COMPAG.2012.08.009.
55. Zhou C, Lin K, Xu D, *et al.* Near infrared computer vision and neuro-fuzzy model-based feeding decision system for fish in aquaculture. Computers and Electronics in Agriculture 2018; 146: 114–124. doi:

10.1016/j.compag.2018.02.006.

56. Zhao Jian. Research on precise feeding of swimming fish in recirculating aquaculture (in Chinese). Hangzhou: Zhejiang University; 2018.

Journal of Autonomous Intelligence (2021) Volume 4 Issue 2.
doi: 10.32629/jai.v4i2.495

Original Research Article

Research on Key Methods of Visual Human-computer Interaction Comfort Quantifacation in Multidimensional Information Space

***Wen Lv*[1]*, *Wenke Kang*[2]**

[1] *Sichuan Institute of industry and information technology, Chengdu 610017, China*

[2] *Northwestern Polytechnical University, Xi'an 710072, China*

ABSTRACT

With the increase of information exchange types and the acceleration of information exchange speed, people not only pursue fast and efficient interaction effect, but also begin to pursue natural and harmonious human-computer interaction comfort experience physically and psychologically. Therefore, human perceived comfort has become an important index in the design of modern human-computer interaction system. However, since comfort is the subjective feeling of human body, it is difficult to realize quantitative measurement and evaluation in the design process, which brings great difficulties to computer intelligent human-computer interaction design and scientific evaluation. Vision is the most important sense of human beings. More than 80% of all kinds of information received by people from the outside world is obtained through vision. Vision is an important means of human-computer interaction in complex multi-dimensional information space, and it is also the most intuitive source and effect embodiment of human comfort perception. Through a large number of experiments, import the test data into the software Origin, and draw the change curve of visual comfort with reading time, which fit with the S-shaped curve as the reference. In this way, the relationship between visual comfort and reading time can be obtained. The mathematical model of the relationship between reading time and visual comfort is established, and experiments are carried out to verify the corresponding mathematical model relationship.

Keywords: *Vision; Human Computer Interaction; Quantification of Comfort; Multi-information Fusion; Matching Model*

ARTICLE INFO

Received: Aug 6, 2021
Accepted: Sep 29, 2021
Available online: Oct 2, 2021

***CORRESPONDING AUTHOR**

Wen Lv
lvsasa@163.com;

CITATION

Lv W, Kang W. Research on key methods of visual human-computer interaction comfort quantification in multi-dimensional information space 2021; 4(2): 24-30. doi: 10.32629/jai.v4i2.495

1. Introduction

With the development of economy and society, natural and harmonious interaction has become the ultimate goal of modern and future human-computer interaction design, and it needs the coordinated participation of human vision, hearing, tactile sensation and other multi-senses. With the rapid development of modern human-computer interaction technology and information system, it requires human-computer interaction system to be more accurate, fast, efficient and comfortable. In order to achieve this goal, in the process of human-computer information interaction, we must give full play to the comprehensive efficiency of human vision, hearing, tactile sensation and other perception methods, and form an all-round multi-information fusion interaction system, so that the operator can fully perceive and master information. At the same time, we should balance the perception ability of various sensory organs, and maximize the perception ability of each sensory organ, so as to improve the comprehensive efficiency of human-computer interaction.

2. Visual human-computer interaction

2.1 Research background and significance

Vision is the most important sense of human beings, and comfort is the subjective feeling of human body. The wide application of computer and network technology in particular makes the accurate measurement and scientific expression of color very important and urgent[1]. At present, the research on comfort is generally qualitative description, and quantitative calculation is very difficult. It is unable to establish the quantitative measurement model between visual human-computer interaction and comfort, complete the demand of real-time quantitative measurement in the process of human-computer interaction, provide scientific and quantitative design reference and evaluation standards, and meet the requirements of intelligent design of computer human-computer interaction in the future. Therefore, we should study the perceptual characteristics of visual human-computer interaction in multi-dimensional information space and the integration mechanism of visual, force tactile, auditory and other perceptual systems, detect the visual comfort quantitatively, and build a multi-dimensional information space human-computer interaction comfort measurement calculation model integrating vision, hearing and tactile sense. These experiments can not only measure the comfort of each perceptual system, but also measure the comprehensive comfort of multiple senses at the same time, which can realize the organic integration of visual, auditory, tactile and other multi-information, and use different methods for different requirements, so as to obtain the results most in line with ergonomics. Improving the efficiency and quality of complex human-computer interaction system and providing consumers with more natural human-computer interaction experience has great theoretical significance and application value.

Mobile reading is a popular way of fragmented time management. People can obtain multi-directional and diversified information by reading news and novels on mobile phones. According to the *2019 China Digital Reading White Paper* (**Figure 1**), by 2019, the total number of Chinese digital reading users had reached 470 million. The data of the 18th National Reading survey released by China Academy of Press and Publication also shows that in 2020, the access rate of digital reading methods (online reading, mobile reading, e-reader reading, pad reading, etc.) was 79.4%, an increase of 0.1 percentage points compared with 79.3% in 2019. It can be seen that people spend a lot of time on mobile reading every day. It can be said that mobile reading has become a part of people's life. However, people often suffer from eyestrain such as dry eyes, sore eyes, dizziness and headache after reading on mobile phones for a long time. The word "ashenopia" is derived from the Greek language, which originally means eye weakness. After using the eyes for a long time, it produces adverse reactions such as orbital pain, blurred vision, eye burning, dry eyes, tears and so on.

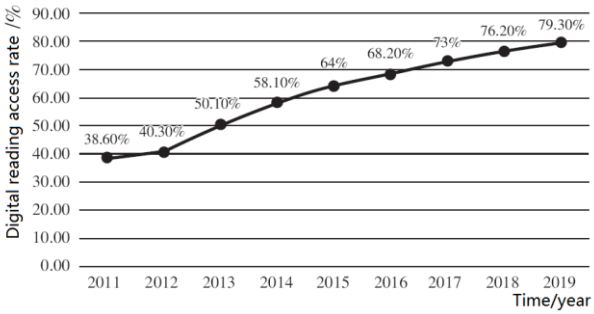

Figure 1. 2019 digital reading white paper.

With the rapid development and popularization of computer technology, mobile phones and tablet computers can also be included in the scope of Visual Display Terminal of personal computer and computing system. VDT syndrome refers to the comprehensive symptoms such as eye discomfort caused by staring at the video terminal for a long time[2]. Visual comfort degree refers to the subjective feeling of the eye when the eye receives external electromagnetic waves such as VDT, which is specifically reflected in the degree of eye fatigue, blinking frequency, visual acuity etc. after using the eye for a period of time. Most of human perception activities are completed through the visual system. At present, the simulation of visual perception is mainly based on the model, in which the sensing information is used to reconstruct the model of the perceived world[3]. This study uses

the methods of theoretical analysis and experimental verification to study the multi-dimensional information perceived comfort based on vision. Taking reading mobile e-books as an example, this study studies the subjects' visual comfort after reading mobile e-books for a period of time. It tracks the relationship between their visual comfort and time under certain conditions, and realizes the quantitative calculation of perceived comfort of visual information. Thus the quantitative model of multi-dimensional information human-computer interaction comfort based on vision is established. Finally, the verification test is carried out to verify the model. The research results can be widely used in complex operating systems, virtual reality and other system environments, and they have important theoretical and application value for improving the level of human-computer interaction design.

2.2 Analysis of research status at home and abroad

At present, the main direction of domestic relevant research is on various factors of electronic display equipment and human visual comfort, such as environmental illumination, font, font size, spacing, and color matching, which is more in line with the visual application scene of people in the development of the times. Most of the research is aimed at some main factors in a specific environment of application. The related achievements include the following. Jiang Ying *et al.* studied and summarized the quantitative model of the visual comfort of the information display interface of the visual display terminal (VDT) through the combination of subjective and objective experiments. The study took the dial interface with different color combinations as the test object, took the physiological characteristics such as EEG signals of the subjects as the objective basis, and supplemented by the subjective score of the subjects on the comfort of the dial interface. The evaluation method is scientific. Hou Guanhua *et al.* studied the influence of font size and spacing on the digital reading experience with different age groups. They comprehensively considered the characteristics of reading experience of groups of different ages, adopted the combination of subjective and objective observation and evaluation, selected the font size and spacing as the research variables, and studied the difference of the impact of different font size and spacing on the reading experience of users of different ages based on the standard of visual comfort. In addition, Liu Chang and others quantitatively studied the influence of chromaticity on the visual comfort of stereo images. Cai Jianqi *et al.* studied the influence of environmental illumination of ipad and other electronic products on human visual comfort in the bedtime use environment. It can be seen that there are few overall quantitative studies on visual comfort at present.

3. Evaluation Index

Visual health and comfort is an objective quantitative index of visual fatigue. It realizes the objective evaluation of everyone's eye health and comfort based on the changes of human eye axis and corneal refraction[4]. The main factors affecting visual comfort include illumination level[5], brightness contrast of screen display[6], light environment (such as playground, reception hall, bedroom, etc.), text size, text interval, eye use time, illumination, color combination, age, etc. When establishing the mathematical model of visual comfort, it must be taken into account that visual comfort has two characteristics—"psychological feeling" and "depends on a variety of uncertain factors". We take the change of reading comfort with reading time as an example to carry out the research. According to the literature analysis, we take the usability and perceived visual comfort as the subjective measurement indicators, directly ask the subjects' feelings during the test, and use the scale to measure the subjects' subjective feelings. This method is the most direct method to evaluate whether the vision is comfortable or not. The research scale is shown in **Table 1**.

Table 1. Limited range of visual comfort and human representation state

Visual comfort value	State
0	No fatigue
1	Basically no fatigue
2	Slight fatigue
3	There is obvious fatigue, but it is within the tolerance range
4	Increased fatigue and various eye discomfort

4. Test Part

The design scheme of repeated measurement test between groups is adopted in the test, and the mobile phones with their own reading app is selected as the test standard. The subjects are 20 volunteers aged 19–22. All of them are college students from Northwestern Polytechnical University, and they don't have severe myopia and astigmatism (myopia within 300 degrees and astigmatism within -100 degrees).

In the study of comfort, fatigue is the most important factor in the psychological factors affecting comfort[7]. It is carried out indoor with normal lighting, quiet and comfortable environment. In order to avoid the interference of visual fatigue, all subjects have enough sleep before the test, and are arranged to close their eyes and rest until there is no fatigue before the test, so as to ensure that the subjects can be in a good psychological and physiological state during the test.

Before the test, we record the basic information of the subjects, and explain the test process to the non subjects. Then we uniformly arrange the subjects to rest until the comfort test is 0 (completely comfortable). Every time the subjects use the mobile phone to read e-books for 5 minutes, we immediately ask the subjects about their subjective visual comfort and fill in the scale. Each participant needs 50–60min to complete all contents. The test process is shown in **Figure 2**.

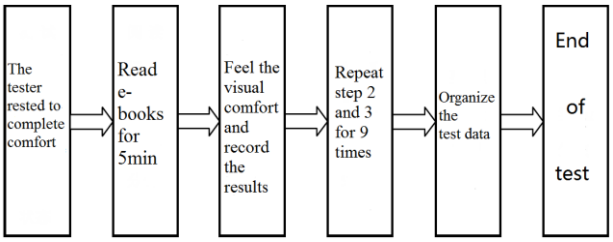

Figure 2. Visual comfort test process.

5. Establishment of Quantitative Mathematical Model of Visual Comfort

Since the study only analyzes the impact of reading time on visual comfort, in order to obtain the mathematical model of visual comfort information perception and the general law between visual comfort and time, it is necessary to ensure that other factors remain unchanged and only record the data of visual comfort at different times. Therefore, the control variable method is adopted in the experiment. 20 subjects were tested for many times in the same time period and the same place, with a total of 20 groups of data. See **Table 2** for the relevant data.

Table 2. Measurement data

Person-time	Time/min									
	0	5	10	15	2025	3035	40	45	50	55
1	0.01	0.01	0.02	0.05	0.140.34	0.781.48	2.19	2.66	2.90	2.99
2	0.01	0.10	0.15	0.25	0.330.64	0.841.50	2.00	2.80	2.90	3.00
3	0.01	0.05	0.05	0.10	0.200.50	0.901.35	1.90	2.50	2.70	2.90
4	0.01	0.02	0.04	0.06	0.120.36	0.721.44	2.20	1.90	1.80	2.00
5	0.01	0.10	0.15	0.18	0.350.70	1.351.80	2.20	2.70	2.85	2.95
6	0.01	0.05	0.10	0.10	0.150.40	0.751.45	2.10	2.55	2.75	2.80
7	0.01	0.02	0.08	0.12	0.160.30	0.701.40	2.15	2.75	2.85	2.95
8	0.01	0.07	0.16	0.21	0.380.62	0.981.38	2.04	2.41	2.68	2.87
9	0.01	0.01	0.06	0.12	0.210.33	0.891.41	1.88	2.34	2.79	2.96
10	0.01	0.02	0.09	0.15	0.310.73	1.051.48	1.95	2.42	2.73	2.84
11	0.01	0.09	0.16	0.31	0.450.68	1.011.42	1.89	2.43	2.82	3.01
12	0.01	0.03	0.05	0.09	0.210.44	0.831.34	1.98	2.56	2.87	2.96
13	0.01	0.08	0.14	0.23	0.310.69	0.981.44	2.08	2.77	2.93	3.05
14	0.01	0.02	0.05	0.17	0.290.49	0.891.35	2.11	2.57	2.72	2.90

Table 2. continued.

15	0.01	0.02	0.04	0.09	0.190.36	0.821.31	2.20	1.97	2.10	2.30
16	0.01	0.07	0.15	0.26	0.350.81	1.251.70	2.12	2.52	2.85	3.15
17	0.01	0.01	0.10	0.13	0.210.40	0.851.52	1.89	2.33	2.55	2.73
18	0.01	0.02	0.04	0.10	0.150.30	0.631.20	2.05	2.75	2.82	3.14
19	0.01	0.01	0.01	0.07	0.170.42	0.961.32	1.93	2.52	2.86	2.97
20	0.01	0.01	0.03	0.08	0.260.45	0.891.56	2.20	2.70	2.86	3.06

For the measured data, take the number of test records as the x-axis and visual comfort as the y-axis, as shown in **Figure 3**, and draw a scatter diagram in the X-Y coordinate system. From **Figure 3**, it can be seen that the changes of visual comfort of these 25 subjects are roughly divided into three stages: stable growth period, rapid growth period and slow growth period. With the increase of time t, visual comfort y almost cease to increase at first, then increases rapidly, and finally slows down. And the distribution of points is closer and closer to a straight line parallel to the x axis, so the y value will not increase infinitely.

According to the change law and trend line of visual comfort in the scatter diagram, it is found that the change trend line of visual comfort and time is approximately similar to the shape of S. The data is taken into Origin nonlinear fitting. Through comparison, it is concluded that when the fitting model is slogisticl1 S type function, the fitting degree is the highest, and the R square is 0.991. At this time, the equation is as the following.

$$y = \frac{a}{1 + e^{-k(x-xc)}} \tag{1}$$

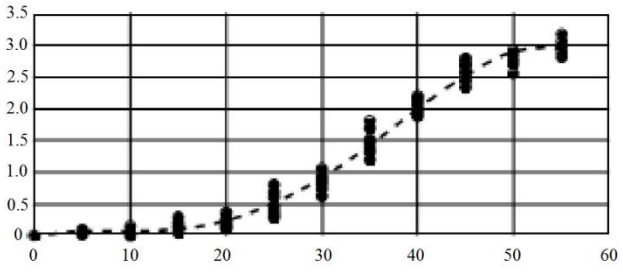

Figure 3. Relationship curve between test time and visual comfort.

The fitting curve is shown in **Figure 4**.

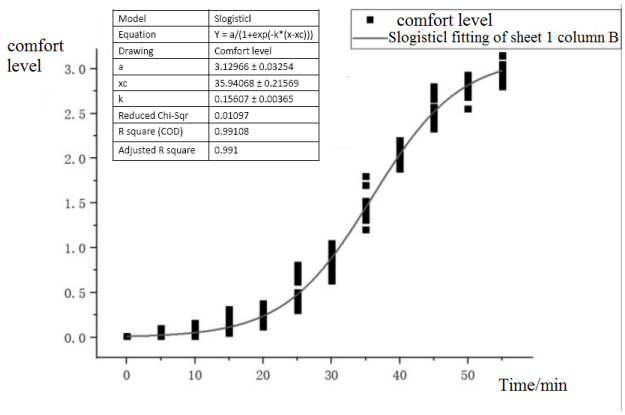

Figure 4. Fitting curve.

The parameters of the fitting equation: a = 3.12966, xc = 35.94068, k = 0.15607.

Therefore, the fitting equation is as the following.

$$y = \frac{3.12966}{1 + e^{-(0.15607\times(x-35.94068))}} \tag{2}$$

Formula (2) is the expression of the relationship between the visual comfort value y in information perception and the time variable x. The quantitative model of visual comfort has been established.

6. Mathematical Model Verification of Comfort Quantification

In order to verify the accuracy of the mathematical model, another five test experimenters were conducted under the same conditions. Compare the comfort degree of test data with the visual comfort degree obtained through the transformation of mathematical model relationship. See **Table 3** for validation data. **Figure 5** is the scatter diagram of the validation data compared with the relational curve. According to the verification data in **Table 3**, the mathematical model of visual comfort obtained from the original data in the test through the calculation of relational expression is basically fitted. Therefore, it can be verified that the whole process of measuring comfort data

through test, analyzing the data, obtaining the mathematical model, and finally calculating the corresponding time through the mathematical model is correct.

Table 3. Model validation data time/min

	0	5	10	15	20	25	30	35	40	45	50	55
1	0.01	0.03	0.05	0.10	0.28	0.52	0.99	1.45	2.10	2.58	2.75	2.98
2	0.01	0.02	0.04	0.06	0.12	0.36	0.72	1.44	2.06	2.46	2.62	2.90
3	0.01	0.05	0.15	0.18	0.35	0.66	0.89	1.50	2.05	2.62	2.75	2.95
4	0.01	0.05	0.15	0.26	0.35	0.70	0.92	1.55	2.10	2.50	2.85	3.15
5	0.01	0.01	0.10	0.13	0.21	0.62	0.85	1.52	1.99	2.62	2.80	3.00
Calculated comfort value	0.011 42	0.024 82	0.053 67	0.114 79	0.240 09	0.480 36	0.887 26	1.45 017	2.044 58	2.517 43	2.815 85	2.977 6

7. Conclusion

A large number of experiments on the test objects are conducted, and the test data are imported into Origin software, drawing the change curve of visual comfort with reading time. The S-shaped curve is used as a reference for fitting, and the relationship between visual comfort and reading time is obtained, establishing the mathematical model of the relationship between reading time and visual comfort. The results show that the mathematical model can better reflect the objective law of visual comfort changing with reading time. The mathematical model can provide a theoretical basis for visual time allocation in design and production, making the product more humane in visual human-computer interaction and more in line with the design concept of "people-oriented". Of course, this part can also be used as a guiding scheme in the process of design and production, so the model owns important practical value.

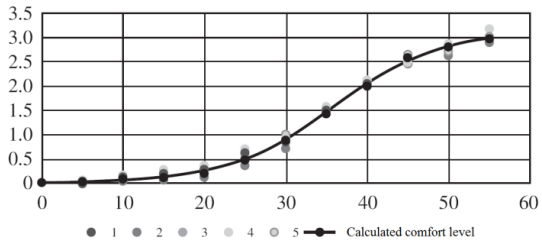

Figure 5. Scatter plot of validation data compared to relational curve.

Due to the complexity of human visual system and the diversity of influencing factors of visual comfort, the quantitative research of visual comfort must be a long-term and systematic process. This paper only studies the impact of reading time on visual comfort. Other influencing factors such as flicker, binocular brightness difference, binocular color difference, and individual physiological factors of human eyes, still need further research and exploration.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Kang W, Zhang D, Yu S. Research on key methods of color conversion of color analyzers (in Chinese). Journal of Environmental Sciences 2010; 46(1): 51–53.
2. Cai J, Du P, Guo Y, *et al*. Research on the influence of ambient illumination on the visual quality of human eyes during the use of ipad (in Chinse). China Illuminating Engineering Journal 2016; 27(6): 6–9.
3. Kang W, Zhang Q, Zhang D. Research on compound thinking in product color intelligent design (in Chinese). Application Research of Computers 2009; 26(6): 2083–2085, 2100.
4. Cai J, Wang W, Shao G. Research on visual health comfort test of LED TV in 3D mode (in Chinese). Television Technology 2014; 38(24): 47–49, 54. doi:10.16280/j.videoe.2014.24.007.
5. Taptagaporn S, Saito S. Visual comfort in VDT operation:physical resting states of the eye. Industrial Health 1993; 31(1): 13–28. doi: 10.2486/indhealth.31.13.

6. Wang AH, Chen MT. Effects of polarity and luminance contrast on visual performance and VDT display quality. International Journal of Industrial Ergonomics 2000; 25(4): 415–421.

7. Luo S. Research on driving comfort based on biological responses (in Chinese). Hangzhou: Zhejiang University; 2005.

Journal of Autonomous Intelligence (2021) Volume 4 Issue 2.
doi: 10.32629/jai.v4i2.497

Review Article

Artificial Intelligence and Human Condition: Opposing Entities or Complementary Forces?

Diego Felipe Arbeláez-Campillo[1]*, Jorge Jesús Villasmil Espinoza[2], Magda Julissa Rojas-Bahamón[2]

[1] *Universidad de la Amazonia, Colombia*

[2] *Universidad del Zulia, Venezuela*

ABSTRACT

In the 21st century, artificial intelligence is a force that surpasses artificial intelligence in many aspects, because it has appeared in all fields of social life, from the Internet search engine that determines the taste and preference of obtaining digital information to the intelligent refrigerator that can issue purchase orders to maintain its availability when some food is exhausted. The purpose of this paper is to analyze the ethical, ontological and legal problems that may arise from the wide use of artificial intelligence in today's society, as a preliminary attempt to solve the problems raised in the title. In terms of methodology, this is a paper prepared using written document sources, such as: literary works, international news articles and arbitration articles published in scientific journals. Its conclusion is that AI may change the lifestyle of the whole civilization in many ways, and even negatively change the human condition by changing human identity and genetic integrity, and weaken people's leading role in building their own realityd.

Keywords: *Artificial Intelligence in the 21st Century; Significance of Human Conditions; Ethical, Ontological; and Legal Conflicts; Complementary Forces*

ARTICLE INFO

Received: Aug 6, 2021
Accepted: Sep 29, 2021
Available online: Oct 9, 2021

***CORRESPONDING AUTHOR**

Diego Felipe Arbeláez-Campillo
dfaca@hotmail.com;

CITATION

Arbeláez-Campillo DF; Villasmil Espinoza JJ; Rojas-Bahamón MJ. Artificial intelligence and human condition: Opposing entities or complementary forces? Journal of Autonomous Intelligence 2021; 4(2): 31-39. doi: 10.32629/jai.v4i1.497

1. Introduction

What is intelligence? Are humans really intelligent? Do humans monopolize intelligence in front of other life forms? Is life limited to organic organisms? In order to contribute assertively to the current debate on the transcendence and importance of artificial intelligence in the 21st century, efforts should be made to answer these questions, emphasizing that the concept of intelligence is by its very semantic essence: polysemic and at the same time multidimensional. Therefore, cognition, emotion, behavior, multiple intelligences and artificial intelligence are all daily topics. Traditionally:

Intelligence is a very common mental ability, including reasoning, planning, problem solving, abstract thinking, understanding complex ideas, rapid learning and learning from experience (...) it requires a broad and profound ability to understand the environment, be able to capture and give meaning to things, or figure out what to do[1].

Ardilla[2] believes that intelligence is the core ability to exist, create and live in different fields as diverse as: Logical–mathematical domain, linguistic skills, musical talent, interpersonal and social interiors, and the systemic bodily dimension. For psychology and neuroscience, it is a measurable phenomenon, and its development or degradation is related to logical, environmental, educational and cultural factors. According to this idea, the concept of artificial intelligence is subsidiary

to the theory and concept of human intelligence to some extent, and its particularity lies in:

(...) specifically refers to a special type of technological intelligence. Although its starting point is human, which is human art and the main reason, it can operate independently and autonomously, and even surpass human cognitive and procedural ability in many aspects[3].

However, the concept of intelligence has been reified by a anthropocentric and logocentric discourse, which implicitly assumes the so-called intellectual superiority of mankind without questioning its paradoxes, but to what extent are the pollical system that have been historically produced and reproduced intelligent ? When they are troubled by the obvious authoritarian contradictions in asymmetric power relations, they subject citizens to the authorization of the government, no matter how absurd they may sometimes be, to what extent the existing economic system intelligent? When they systematically plunder non-renewable natural resources for profit and concentrate their wealth in 1% of the population of developed countries at an exponential rate[4].

Fortunately, AI can overcome the limitations and contradictions of human intelligence in many aspects, deepen its position as a complementary force of intelligence, or on the contrary, eventually lead to hostile factors, which will be proved in the future. It is these and other similar ideas that drive the development of this article. It is worth mentioning that in many literary or film works such as Isaac Asimov's *I, Robot* or James Cameron's *Terminator*, AI contacts humans in the form of robotic-anthropomorphic to the same extent as the high degree of autonomy, enabling these" entities "to deploy a set of decisions that may be ethically, ontologically or legally controversial.

However, there is no doubt that in the 21st century, artificial intelligence goes beyond action in many aspects, because it exists in various fields of modern social life in some way, from Internet search engines to determine tastes and preferences for obtaining digital information, to smart refrigerators capable of issuing purchase orders to maintain availability of some foods as they run out. Therefore, the purpose of this paper is to analyze the potential ethical, ontological and legal problems caused by the wide use of artificial intelligence in today's society, as an attempt to solve the problems raised in the title, and as an excuse for debate on this important issue.

In addition, this paper is divided into four separate but logically related parts. In the state of art, the selected works that shaped the concept of artificial intelligence and its different effects on human conditions are presented. The second section explains the methodological procedures for interpreting and organizing consulting sources, and clarifying the cognitive positions of researchers. In the third section, there is an interest in addressing the proposed objectives, at least temporarily, in order to arrive at conclusions and findings of the case in the last part.

2. State of art

According to the limits of the length of the scientific article genre, we briefly review different works and authors whose contributions underpinned the theoretical and analytical apparatus of this article, and also served as an influence to structure this vision of artificial intelligence in today's world. This is a technical phenomenon that under appropriate circumstances, can make a significant contribution to the transformation of the whole human civilization in unexpected ways.

The work of Vinuesa *et al*.[5] was important to visualize the practicability and versatility of artificial intelligence in modern society. In the words of the author, artificial intelligence plays an important role in promoting sustainable development goals through the implementation of algorithms and software. These algorithms and software can supervise the realization of more than 100 goals with the support of qualitative and quantitative information. However, the current limitations of this form of autonomous intelligence are not excluded, highlighting some gaps in transparent data processing, security vulnerabilities and the proper implementation of ethical standards.

Similarly, the research of André and Romy[6] also helps to clarify the preconditions for integrating artificial intelligence into the teaching process of higher education and to study the new business model of *Ed Tech* technology. The author believes that with the development of social reality, new business models have

emerged, which have an impact on the market and education.

Therefore, a considerable number of educational technology companies (*Ed Tech*) are trying to update the traditional educational model through the systematic implementation of artificial intelligence to deal with different types of metadata, which come from the daily use of web search engines and other sources reflecting people's tastes and preferences. They concluded that uncertainty and inadequate understanding of strategic data are hindering the development of solutions, so artificial intelligence is the best tool to promote the development of such businesses.

For their part, Belk, Humayun and Gopaldas[7] pointed out that AI is not necessarily a typical novel concept of Western modernity, because some ancient civilizations tried to represent AI like concepts through magic and religious activities such as alchemy, thus establishing a mythical tradition in which humans are fascinated or afraid of another form of alternative intelligence to challenge their intellectual hegemony on the planet or surpass their cognitive and procedural abilities.

In conclusion, Belk *et al*.[7] pointed out that artificial intelligence methods based on software (soft) or robot (hard) can be expressed in humans, resulting in a difficult process to determine a priori of transhumanism. There is an urgent need to establish a framework for human dignity and social security for all practical reasons. Finally, two discourse traditions originated from artificial intelligence are valued: on the one hand, skeptical and cautious views, on the other hand, futuristic views welcome this form of intellectual progress without further doubts.

Biagini *et al*.[8] also emphasized the autonomy of artificial intelligence in managing the power system in the near future, which is a possible condition for establishing a sustainable energy park with little human supervision. The author believes that the introduction of AI software and hardware can also promote the technological evolution transition of human beings to a higher life cycle stage, with AI produces and reasonably distributes abundant renewable energy.

Of particular interest to the research team were the works of the controversial Hebrew historian Yuval Noah Harari[9, 10] which now comes to confirm with specific empirical evidence and makes predictions of science at different times, even surpassing them in many ways. For him, intelligent people are on the verge of extinction. Due to the combination of genetic engineering, robotics and artificial intelligence, another post-human or superhuman entity (as seen in the phenomenon) will be replaced by another biomechanical and cybernetic entity, which will overcome the material and cognitive limitations of human conditions and develop their skills and abilities at an exponential speed. In this regard, he believes that:

Cyborg engineering will go a step further by integrating organic bodies with non-organic devices, such as bionic hands, artificial eyes or millions of nanorobots, which will navigate our bloodstream, diagnose problems and repair damage. This type of cyborg will be able to enjoy far more power than any organic object[10].

It also emphasizes that:

(...) Cyborg can exist in many places at the same time. Dr cyborg can perform emergency surgery in Tokyo, in Chicago and on a space station on Mars without leaving his office in Stockholm. All she would need is a fast Internet connection, a few pairs of bionic eyes and hands. But think about it, why pairs? Why not quartets? In fact, even these are really superfluous. Why would a cyborg doctor have to hold a surgeon's scalpel when he could connect his mind directly to the instrument?[10]

There is no doubt that this information may be exaggerated and fantastic from a conservative point of view, but even an superficial review of the history of the last century shows that the dazzling scientific progress has achieved unimaginable achievements in daily life for generations, such as real-time communication with people all over the world, provided by intelligent mobile devices with wireless broadband Internet; the open socialization of scientific knowledge or the promotion of new forms of work, interpersonal relationships and virtual education facilitated by Information and Communication Technology (ICT). Therefore, in the coming decades, Harari's avant-garde thought today will certainly be replaced by the future.

3. Clarification of epistemology and methodology

The authors of this work believe that mankind will soon undergo a transformation of the world order[11, 12], partly due to the devastating impact of the new coronavirus and the depletion of the political and economic models that have worked so far.

It is reasonable to assume that these upcoming changes will, in turn, promote the paradigm shift of scientific structure, consolidating the new rationality described by Martinez[13], so as to break through the restrictions imposed by neopositivism. At least in the field of social and human sciences, as a fruitful attempt to re-examine humanistic knowledge, such as philosophy, literature and art, without prejudice, within the framework of a unified systematic conception of scientific that, without losing its rigorousness in its methodical and theoretical bodies, is simultaneously : objective-subjective, inductive but deeply interdisciplinary; perhaps recalling the integrity and holistic sense of integrity of his understanding of the ancients or people in the Renaissance.

It is inspired by these Gnostic gestures that herald the paradigm revolution that prose can appear effectively, not just as a breakthrough of a literary school. It emphasizes:

(...) Because of its freedom of form and theme, and because of its ideology-although it appears mainly as a hostile response to a given gender-this article is a typical example of the visibility of general issues. From Aristotle to today, any position of literary thought is on the pillar of genre theory, which is a limit, if not a blind spot[14].

In fact, critical writing is also a key tool of general scientific discourse. It allows the course of a free hermeneutics to be expounded with some originality under the combination of the following factors: the author's creativity, criticality and flexibility in dealing with sources go beyond the rigid formal standards of traditional monographs. These monographs reproduce that cognition belongs to the dominant paradigm and obliterate innovation to some extent.

In terms of operation, the theme was first identified to recognize the importance of artificial intelligence to today's and future world. Subsequently, a large number of written literature sources were recorded in physical and digital form, including scientific articles in recognized databases: *Scopus,* WOS, Dialnet, etc; in addition to text and gender differences, international literary and news articles agree that artificial intelligence has an unusual central position in its various forms and ways of existence. Finally, the work was drafted in accordance with the authors' rules of the Journal of Social Sciences for evaluation, comparison and publication if appropriate.

4. Artificial intelligence and human condition: Ethical, ontological and legal issues arising from its widespread use in society

When discussing the great possibility of the wide spread of artificial intelligence in society, we can quickly see at least two views: on the one hand, it reveals that individuals and researchers believe that this form of intelligence is a multidimensional tool to improve human living conditions. They promote work, process and decision-making in an autonomous and equal manner according to the interests and needs of people in different backgrounds. On the other hand, there are also voices of criticism, which indicates that artificial intelligence may completely change the historical process of society in the 21st century incurring in anachronism or retrogression of the times.

On this main line, without falling into the fallacy of wrong dilemma, every requirement on this issue means a certain positioning of the pointed position, and this article is no exception. To date, however, AI has shown significant versatility in helping manage almost all human activities, such as: education, health, finance, entertainment, housekeeping, and even engage in some free professions dominated by technology, intellectual knowledge, creativity and experience. For example, the apps available on smartphones today include: ADA, *iDoctus* and IA based Endomodo can make accurate medical diagnosis of various diseases, while other diseases such as: *E Law Guide, SmartLeges* does the same thing on legal issues.

However, scientists like Stephen Hawking[15] did not hesitate to predict the danger of artificial intelligence to humans, saying: "Human beings, limited by their slow biological evolution, will not be able to com-

pete with machines and will be surpassed" (p. 1). According to the late British physicist, a complete artificial intelligence can also be self-designed. Its purpose is to overcome all artificial control and affect people according to its rapid development, without denying the potential of this intelligence in solving various problems[15].

Anticipating this situation, robotics has become the largest material manifestation of artificial intelligence and human conditions. The famous scientific writer Isaac Asimov has long formulated three basic rules of robotics:(1) Robots will not harm humans, nor will they let humans be harmed indifferently. (2). Robots must issue or execute commands issued by humans unless they comply with the first law. (3). Robots must protect their existence as long as this protection does not comply with the first or second law[16].

From the reading between the lines of these rules or laws in the short story *Runaround* published in 1942, it can be seen that the fate of every robot is to subordinate its action to the orders of human beings, so as to ensure that under the first clear premise, people will not suffer any damage due to their own existence.

In addition, it is obvious that Asimov proposed the beginning of the ethical debate. He tried to standardize the practice of artificial intelligence regulating from the field of literature, so as to avoid any contact or tension between these new entities and the whole mankind. In fact, this ethical concern is prominent in his main works: *I, robot*; *Bicentennial Man* and the *Foundation Sega*[2].

There is no doubt that artificial intelligence is vulnerable to ethical treatment in all its manifestations and modes, because its own risks and possibilities implied by its own existence, which are constantly developing. More importantly, all signs show that it is an entity with dialectical development and has the ability to act, learn and shape itself at a speed beyond human ability according to its own personal and collective experience. In fact, selected circles such as the Bilderberg Club have included the AI paradigm as a force for change in modern society on their international discussion agenda[17], which has produced a series of conspiracy theories on social media.

On this key point, it is worth clarifying that ethical reflection, according to Savart[18], is the philosophical concern that involves the exercise of individual freedom that opts for the pursuit of a good life, all of which are from a moral point of view. Similarly, the view of the problems arising from the exercise of freedom is a subsidiary part of Fromm's[19] humanistic ethics. For Fromm, any ethical requirements are based on a formal and material standard, which connects individual practice with general or specific, utilitarian or dogmatic moral norms, based on the principle that only human conscience can determine the quality of their behavior, not necessarily an authority beyond it, that is, to promote the self-determination of will without the need for external coercive forces such as state or religion.

In both cases, Savat[18] and Fromm[19] implicitly believe that man is the only person endowed with conscience and will in essence and existence, and can act freely, so as to defend ethical treatment. Although these anthropocentric positions have dominated matter so far, the development of artificial intelligence challenges that intelligence is a unique attribute of mankind, and life is limited to the dominant paradigm of its biological essence, which at least theoretically allows a more meaningful post-human ethical debate in the coming decades, just as all doubts about the special will and consciousness of artificial intelligence have been eliminated, they exist and manufacture according to the conditions of hardware and software, which determine their own existence and self-driving force.

Without delay, what are the main ethical issues that may arise directly or indirectly from the spread of artificial intelligence to all aspects of social life from now on? In this regard, it should be remembered that, just as the first industrial revolution caused massive unemployment by replacing human beings with cheaper and more efficient machines at that time, in the upcoming revolution of telematics and nanorobots dominated by artificial intelligence, the situation will not be much different from that in Age of Enlightenment, and the results in this regard will be more remarkable, because these entities have multiple uses in almost every conceivable field.

In the case described, as Duseel[20] understood, this is an event that substantially changes the principle of political *ratio,* because a new world order dominated by artificial intelligence, as a material practical

rationality, will not only solve the conditions that decide to create and reproduce a better life for mankind, but also introduces an external force into the key process of maintaining the balance of political and social ecosystem, such as the production of goods and services, the distribution of value goods and the division of social labor and knowledge.

One might argue that the impact of artificial intelligence will lead to reduced working hours for individuals and communities as a whole, while increasing space for entertainment, family life and training, and that the loss of potential jobs selected by intelligent machines will be compensated by public policies that ensure the general basic income of all, similar to the social security experience of the welfare state after World War II. However, all historical evidence shows that in the global south, these policies have been inadequate and limited. Under no circumstances can they guarantee the resources needed to get rid of poverty and unstable life. The opposite argument is either illusory or demagogic.

If this is added, the main interests of AI may be monopolized by political and economic elites, as happened in Latin America with various modernization plans. According to Morales, Villasmir and Martinez[21], even today, these plans cannot provide positive solutions to the poverty, inequality and the lack of opportunities for the vast majority of people in the region to support their development, Therefore, AI will increase inequality and the mechanisms of social exclusion, because it will provide a series of comparative advantages for a few people and harm those who cannot enjoy these technologies.

In addition, AI will further strengthen the formal and informal social control mechanisms of authoritarian governments to destroy the space for democratic participation and citizen leadership, which has been achieved through arduous struggles all over the world in history. If this claim is exaggerated, we just need to see how the social network called Social Credit System has worked so far in China. The system combines facial recognition, real-time geographic location and artificial intelligence software, and uses a scorecard to reward or punish individuals. In some cases, if the "ideological purity of the system" is violated, it will lead to the prohibition on leaving the country, travelling on train or opting for a bank loan[22].

As a representation of ethical issues, a series of ontological issues have also been visualized that it is interesting to review roughly. In this case, ontology is to seek the ultimate meaning of human internal dignity. From this point of view, the existence of human beings or the establishment of their own conditions of existence in history is not only due to the absolute necessity of their biology, but also especially due to the intersubjective action of attributes such as intelligence, will, cooperative ability, etc. that characterize the species and culture, as the force of continuous reorganization of nature.

In fact, the main ontological problems arising from the wide use of artificial intelligence in today's world are those adjusted through its main role in order to transform human beings into a different entity in the future. At this time, the basic characteristics and spirit of these *entities* cannot be fully determined. If Harari's[10] forward-looking view is correct, human species have experienced this phenomenon since the biological and material sunset, which brings more questions than answers: Has the integration of artificial intelligence and human beings subvert the human situation? What are the ethical boundaries of any evolutionary process? Is the new self-generated by the combination of genetic engineering, nano thrombosis and artificial intelligence necessarily unacceptable? As early as the end of the 19th century, Friedrich Nietzsche, a great German thinker, once said:

Man exists only to be transcended. "Man is a rope between beast and Superman-a rope across the abyss." Therefore, its greatness lies in the fact that "it is a bridge, not a goal" and that what is to be loved in it is "it is a transition and sunset"[23].

Perhaps the development of technology, especially the development of artificial intelligence, will be the driving force of human transcendence known in ontology so far. Alternatively, this overall transcendence of *Homo sapiens* may coincide with Nietzsche's transformation from beast to Superman. By the way, at a complex intersection, people are discussing the sunset of human beings or the evolution of their abilities to God's own level, as Harari imagined[9].

It is true that nothing is certain and can only be guessed, but in any case, the changes in the next few decades will not be easy and will bring them with

changes in people's identity and may transform their genetic integrity, significantly alter their being, doing and living together with the way they feel the world and live their life. At this stage of the development of human history, it has always been bound by the limitations of human conditions and subjectivity.

Last but not least, AI needs to replicate a set of laws at any time. These laws test the ability of legal knowledge to regulate and regulate the design, production and use of AI at present and in the future. In this regard, Leret[24] emphasized the close relationship between biotechnology, law and bioethics due to the rapid development of technology in the fields of Informatics, medicine and genetics to protect the identity of individuals and their genetic integrity and identity.

In this regard, there are some reasonable problems, such as: Is a person's gene designed to achieve certain genotypic and phenotypic attributes contrary to the natural evolution of the species? Will these advances will divide humans into two categories: "normal" people and those who are gifted through genetic manipulation and/or biomechanical implantation? Although there is no clear answer in this regard, it is clear that developed countries will be the biggest beneficiaries of these improvements, which will further strengthen the technological and scientific gap between the North and the South.

Since Calvano[25, 26] proposed the political modernity scheme, there has been a concept of justice, democracy and fairness in the western legal system through symbolic and conceptual means, such as: The rule of law and human dignity are designed to protect ordinary people from situations that violate their spirit or worsen their living conditions. In this regard, the establishment of regulators for the regulation of AI should focus on at least four key aspects, which have been regarded as:

Develop in an international legislative framework by consensus to subordinate the design, production and development of artificial intelligence to the dignity of life in its biological type. As postulated in Asimov's first law of robotics.

Identify universal ethical and axiological principles to make AI meaningful and purposeful as a tool for the development of human capabilities, as Nussbaum[27] understands, that is, the immeasurable possibility of existence and realization as part of a life project, which is formulated by individual sovereignty.

Penalize the instrumentalization of human conditions by incorporating of artificial intelligence technological implants which change people's consciousness and/or ignore the ethical standards of people's free existence and action.

Punishment uses artificial intelligence as a formal and informal means of social control to support totalitarian order, in the style of dystopian fictions such as in *Black Mirror*, Aldous Huxley's *Brave New World*, or George Orwell's *1984*.

The success of such a legislative agenda will depend not only on the political will of advanced governments to make full use of artificial intelligence, but also on the pressure of organized civil society to defend their rights in a substantive democracy, in which artificial intelligence will gradually become an unacceptable means of governance, no matter how possible its performance is on various technical equipment.

4. Conclusion

Are artificial intelligence and human state opposite entities or complementary forces? In essence, all this will depend on people's universal use of this technology, its design purpose, and the specific results obtained from it for the benefit or damage of life in general. Of course, whether humans can control this form of intelligence under some ethical and bioethical standards in the triangle that combines artificial intelligence, robotics and genetic engineering. So far, artificial intelligence seems to be able to overcome the limitations and contradictions of human intelligence in many aspects and deepen its position as a complementary force of human intelligence.

However, it should not be ruled out in advance that at a certain stage of its evolution, artificial intelligence contacts human beings in different ways, just as it has developed a high degree of autonomy, enabling it to make a series of decisions that may be ethically, ontologically or legally controversial. Whether due to their own conscious will or their use as a tool for social control and rule, as has happened in countries like China, where artificial intelligence serves dictatorship.

For the reasons described in the full text, two

views on artificial intelligence are considered: on the one hand, as a tool that helps to improve the future of individuals and communities, on the other hand, as a force, it must be critically analyzed through the various threats generated by its direct or indirect use. This choice tends to the second position without compromising its multiple contributions to modern life. All indications indicate that artificial intelligence will play a central role in the new world order being established, which is likely to indicate the transition of mankind to a higher quality entity, or, on the contrary, the moral erosion of human conditions. From this perspective, there are more questions than answers, and there are good reasons to be vigilant about what is about to happen in the global transition stage.

In addition, artificial intelligence is such a broad and complex problem that demands at all times an interdisciplinary research that addresses it in its various dimensions, as a condition of possibility to understand this phenomenon in its dialectical totality. To this end, it is not surprising to propose more and better research routes and take into account the ethical, ontological, legal, political, psychological and anthropological problems arising from the increasingly widespread use of this technology in theory and reality.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Sanz C. What is intelligence? eMS El Mundo del Superdotado; 2021.
2. R. Inteligencia. Qué sabemos y qué nos falta por investigar?(Spanish)[Intelligence. What do we know and what do we need to in-dress?] Rev. Acad. Colomb. Cienc. 2011; 35(134): 97-103.
3. Brito P, Villavicencio C, Sánchez P. Requirements for possible links between artificial intelligence and the future of society. Revista de la Universidad del Zulia 2019; 10(28): 260–280.
4. Stiglitz JE. The price of inequality: The 1% of the population has what the 99% needs. Taurus; 2012.
5. Vinuesa R, Azizpour H, Leite I, *et al*. The role of artificial intelligence in achieving the sustainable development goals. Nature Communications 2020; 11(1): 1–10.
6. Renz A, Hilbig R. Prerequisites for artificial intelligence in further education: Identification of drivers, barriers, and business models of educational technology companies. International Journal of Educational Technology in Higher Education 2020; 17(1): 1–21.
7. Belk R, Humayun M, Gopaldas A. Artificial Life. Journal of Macromarketing 2020; 40(2): 221–236.
8. Biagini V, Subasic M, Oudalov A, *et al*. The autonomous grid: Automation, intelligence and the future of power systems. Energy Research and Social Science 2020; 65: 122–139.
9. Harari YN. From animals to gods: Brief history of mankind. Debate; 2014
10. Harari YN. Homosexuals: A brief history of tomorrow. Debate; 2015
11. Arbeláez-Campillo DF, Andreyevna Dudareva M, Rojas-Bahamón MJ. In a globalized world, epidemics are a factor undermining the geopolitical order. Cuestiones Políticas 2019; 36(63): 134–150.
12. Villasmil J. The fragility of human civilization. Cuestiones Políticas 2020; 37(64): 10–14.
13. Martinez M. New science: Challenges, logic and methods. Thrills; 2009.
14. Camproubi CB. Genre theory research. Estudios de Literatura 2014; (05): 101–123.
15. Stephen Hawking: Artificial Intelligence Could Spell End of Human Race [Internet]. BBC News; 2014. Available from: https://www.bbc.com/mundo/ultimas_noticias/2014/12/141202_ultnot_hawking_inteligencia_artificial_riesgo_humanidad_egn
16. Asimov I. Runaround. New York: Street and Smith; 1942.
17. Esparza R. Bilderberg club has launched a new paradigm: Artificial intelligence [Internet]. Hispanic; 2015. Available from: https://www.hispanidad.com/confidencial/el-club-bilderberg-lanza-un-nuevo-paradigma-la-inteligencia-artificial_212299_102.html
18. Savater F. Amado's morality. Barcelona: Ariel; 1999.
19. Fromm E. Ethics and psychoanalysis. Fondo de Cultura Económica; 2003.
20. Dussel E. Towards critical political philosophy-Human rights and development. Bilbao: Desclée; 2001.

21. Morales Castro YR, Villasmil Espinoza J, Martínez Pure RD. Democracy in the spotlight: The people's government or the degeneration of power? Revista de Ciencias Sociales 2019; 25(1): 236–252.
22. Garrido L, Shanghai J. Less freedom vs. more economic development. Good citizen, force: China has accelerated its social control plan for 2020 [Internet]. El Condoncial; 2019. Available from: https://www.elconfidencial.com/mundo/2019-09-20/buen-ciudadano-a-la-fuerza-china-acelera-su-plan-de-control-social-para-2020_2201731/.
23. Nietzsche F. Immortal works II Volume 2. Tecnos Editorial, S.A; 2003.
24. Leret MG. Law, biotechnology and bioethics. Ediciones CEC; 2005
25. Calvano L. Political Modernity: Gaze and sign. Revista Cuestiones Politícas 2017; 33(58); 37–48.
26. Calvano L. Social contract and political modernity in Colombia. Fondo Editorial de la Universidad Rafael María Baralt (UNERMB) 2018.
27. Nussbaum M. Creating capabilities: a proposal for human development. Editorial Paidós; 2012.

Notes

1. As Ritzer (1999) said: "Redistribution can be seen as a process through which people begin to believe that the social forms created by human beings are natural, universal and absolute. Therefore, these social forms do have these characteristics. The concept of "redistribution" means that people believe that the social structure is beyond their control and cannot be changed "(p. 191). Although one might say that human intelligence has a neurobiological basis, it is not a pure human attribute. We understand intelligence as a basic social concept that changes over time and culture. In addition, all life forms have a certain intelligence, which depends on the characteristics of their species, so that they can adapt to the environment, even in extreme environments, such as the so-called extreme creatures. Therefore, we assume that the concept of intelligence has been redefined.
2. For a review of Asimov's main works, it is suggested to refer to: https://www.europapress.es/cultura/libros-00132/noticia-24-anos-isaac-asimov-10-novelas fundamentales-20160406165615.html.

Journal of Autonomous Intelligence (2021) Volume 4 Issue 2.
doi: 10.32629/jai.v4i2.500

Original Research Article

Application of Artificial Intelligence in Monitoring the Use of Protective Masks

Alexandre Pereira Junior[1], Thiago Pedro Donadon Homem[1], Fabio Oliveira Teixeir[1]

[1] Federal Institute of Education, Science and Technology of São Paulo, São Paulo 3550308, Brazil

ABSTRACT

In the context of current epidemic diseases, this study developed a web application, which can monitor the use of protective masks in public environments. Using the Flask framework in Python language, the application has a control panel to help visualize the obtained data. In the detection process, Haar Cascade algorithm is used to classify faces with and without protective masks. Therefore, the web applications are lightweight, allowing the detection and storage of images captured in the cloud and thte possibility of further data analysis. The classifier presents precision, reversal and f-score of 63%, 93% and 75%, respectively. Although the accuracy is satisfactory, new experiments will be carried out to explore new computer vision technologies, such as the use of deep learning.

Keywords: *Computer application, COVID-19, Facial detection, Haar Cascade, Artificial intelligence, Prevention*

ARTICLE INFO

Received: Aug 18, 2021
Accepted: Oct 10, 2021
Available online: Oct 17, 2021

***CORRESPONDING AUTHOR**

Alexandre Pereira Junior
a. Junior@aluno.ifsp.edu.br;
Thiago Pedro Donadon Homem
thiagohomem@ifsp.edu.br

CITATION

Junior AP, Homem TPD, Teixeir FO. Application of Artificial Intelligence in Monitoring the Use of Protective Masks. Journal of Autonomous Intelligence 2021; 4(2): 40-51. doi: 10.32629/jai.v4i2.500

1. Introduction

At the end of 2019, Wuhan recorded the first undetected disease case in the world. Affected people presented cough, sneezing, fever, and the worst case is severe pneumonia, leading to acute respiratory failure[1]. The patient's clinical condition deteriorated rapidly and the intensive care unit was required to make efforts to restore health. In just a few days, Wuhan has become the main epicenter of the new virus, and the world continues to strive to control the epidemic, such as building hospitals in a record time and strictly implementing social distance measures. An example of these measures is , which aims to limit personal contact with infected individuals, which is necessary for infectious diseases without vaccines or treatment. Therefore, Wuhan was isolated to curb the spread of the virus.

The new virus was named SARS-CoV-2. It has been identified as part of the coronavirus family, the pathogen of COVID-19, a new disease. The spread of this disease around the world has led to a public health emergency of international significance, which is described by WHO as an epidemic. Due to the high transmission rate of the virus, isolation measures need to be taken; It occurs from one person to another by touching, dripping saliva, sneezing, coughing, phlegm, contaminated objects or surfaces. According to the World Health Organization (WHO), basic personal hygiene care can reduce the transmission of the virus, such as hand hygiene with alcohol and gel, and the use of personal protective equipment, such as surgical masks (N95) or handmade masks.

Studies have shown that masks can block 95.15% to 99.98% of aerosols[2].

In addition, the mortality rate of COVID-19 is about 3.4%, which is 9.6% and 34.3% respectively compared with other coronaviruses such as SARS and SRM[3]. In some cases, the clinical development of the virus requires the use of ventilator supported intensive care units (ICU). Due to the speed of disease transmission, some countries face difficulties in meeting the demand for these devices and avoiding as many deaths as possible. Therefore, social isolation and personal protection measures have been essential to reduce the number of infections and avoid overburdening the health system. Countries that postponed such measures, such as Italy and Spain, faced the problem of hospital overcrowding, resulting in thousands of cases and deaths due to the lack of beds, and finally have to decide who would receive appropriate care and who would not.

In Brazil, the first action taken to curb the spread of the disease was implemented at the airports to prevent the entry of people with specific symptoms of the disease, such as cough, fever, runny nose, sore throat, respiratory distress, loss of smell (anosmia), taste disturbance (ageustia), gastrointestinal diseases (nausea, vomiting, diarrhoea), tiredness (asthenia), decreased appetite (hyperoxia) and dyspnoea (shortness of breath). However, the entry of the virus is inevitable, because at some stage of the disease, the infected person is asymptomatic.

Due to population density and human trafficking, the state of São Paulo has become the epiccenter in the country. The city has the busiest airport in South America, providing air bridges for various flights. After confirming the first batch of COVID-19 cases, the state announced quarantine. Educational institutions and non-core activities have been prevented from continuing to operate and are not expected to reopen.

The largest city in the state of São Paulo, according to Decree 59396[4], everyone must wear masks in public places, shops and public transport, otherwise they will be fined. Masks have become an essential element for São Paulo residents to move around the city in an emergency. Because even with the efforts of the government of São Paulo, the population segregation rate is very low; in the period from April to May 2020, this proportion is between 46% and 59%[5], ideally 70%.

Data to measure quarantine compliance are collected through a system that maps population mobility indices, known as the intelligent monitoring system of Sao Paulo[5]. The system is established through public-private partnership and involves national mobile operators. Although SIMI-SP is effective in identifying aggregates and measuring isolation rates, it does not monitor public behavior when crossing public places, such as the use of protective masks.

In order to complement the actions of SIMI-SP, it is necessary to develop a monitoring system to determine whether individuals use personal protective equipment in public places where people circulate. Action of this nature is necessary until an effective drug or vaccine COVID-19 is found.

The use of artificial intelligence (AI) can monitor the behavior of people in transit during a pandemic. According to Pontes[6], AI is a computing field aimed at stimulating problem-solving and decision-making capabilities. According to Backes and Junior[7], their sub field "computer vision is a research field that attempts to transfer amazing visual ability to machines", because computers interpret data differently from humans. Artificial vision focuses on [...] Capture and enhance the image (for example, eliminate noise, increase contrast, etc.), separate the region or object of interest from the scene, extract various information such as shape, color and texture from the analyzed image, and finally associate the image with other previous views.

An improved monitoring method of artificial intelligence is facial recognition. It is used in many fields, from the most aesthetic — for example, to improve photos using the face as the main element in applications such as Instagram. — to other areas such as agriculture and police skills[8]. Face recognition can be used to monitor a large number of people at the same time, highlight their faces, and use algorithms to convert these images into data that can be understood by computer systems.

1.1 Ask questions

Due to the high infectivity of SARS-CoV-2, the use of protective masks is mandatory in several countries as an effective and low-cost alternative to curb the spread of the virus, because it protects the individuals who use it and the people around them. According to the study, even in the case of reduced cases, wearing masks and other measures, such as social distance, may continue until 2022[9]. Therefore, autonomous social monitoring solutions are needed to promote the recovery of economic and social activities by ensuring that measures continue to be taken during and after the most critical period of the pandemic.

Thus, the main purpose of this study is to develop a composite solution to monitor the use of protective masks in public and private environments from the images taken by the camera by using AI and computer vision technology. To that end, the following specific objectives are proposed: 1) Using artificial vision and AI algorithm to build a face detection system; 2) Develop a database to store the obtained data; 3) create a dashboard to monitor the system data.

2. Feasibility study

This section describes initiatives and studies related to the proposed project. A computational solution is proposed to minimize the pan demia impact of COVID-19.

2.1 Related work

Even without face detection technology, it is worth monitoring SIMI-SP[10] monitoring system currently used by the state government of São Paulo. The system produces a map of urban agglomerations and can monitor the rate of social isolation in urban centers and places most conducive to the spread of new cases. It was created by a public-private partnership involving mobile operators. However, this solution cannot determine whether a person is using personal protective equipment, such as protective masks recommended by who and mandatory in some cities.

In September 2020, the "face recognition" descriptor was searched in the virtual science article Library of Web of Science, and 10325 times in the past five years. In this result, 5737 articles were published in scientific journals, 4371 in scientific conferences, and 217 in other ways. Only 116 articles (about 1.1%) were written by Brazilian authors. Although they account for only a small part of the research, some Brazilian institutions have discussed the use of face recognition in their projects.

One of the publications discusses the gender classification of people in the image, identifying whether a person is male or female in real time in an uncontrolled environment through their physical characteristics, and factors that may impair the accuracy of the image (e.g. Site brightness). In order to solve this problem, the author simulated human understanding of gender recognition and established a neural network to recognize these patterns[11].

Another publication of Brazilian scientists is the Deepfake study[12]. This technology uses AI to create false videos, which may lead to false positives in face recognition systems. By using convolutional neural network (CNN), the security system can benefit from the solution proposed in this paper.

Among the selected articles, a Brazilian study also compared traditional machine learning and deep learning methods[13]. According to this article, the model created from the trend database performs better in face recognition task.

2.2 Reason

In view of the above, in the case of a high-risk pandemic, it is necessary to monitor the use of masks by the public in the public environment to ensure the effectiveness of measures to prevent covid-19 infection. This article describes the technology used to build a solution for autonomously monitoring the use of personal protective masks by people in an uncontrolled environment.

3. Architecture solutions

The following describes the architecture of the solution proposed in this project and the relationship between the computing components used.

3.1 Algorithm classifier

The classifier used in this application is based on the technology proposed by Viola and Jones[13] and the *Haar Cascade* method[14], which has been specially used for face, human and other aspects so far. However, the classification process requires image preprocessing.

The image in the computer system is a three-dimensional matrix composed of color, width and height. In the image processing step, they define their default height and width so that they all have the same pixel size. In the process of image processing, some noise can still be removed from the image obtained in the capture process, such as smoothing and highlight filtering. In this work, we decided to omit the dimension of color and convert the image to gray. In this way, the classifier will execute the function in a two-dimensional array.

Haar Cascade method[15] includes three basic steps: 1) track the image and find the Haar features similar to the bus object; 2) use boosting classifier to select the most relevant features; 3) cascade classifiers to improve the final result.

Therefore, in the first stage, it scans the image by rotating small particles from left to right and from top to bottom to find Hal's characteristics[16]. **Figure 1** illustrates the type of manual core used to detect Haar Cascade: image (a) reference core to identify edge features; image (b) reference texture is used to identify the straight line feature, and the image (c) reference texture is used as four rectangles to identify the diagonal edge.

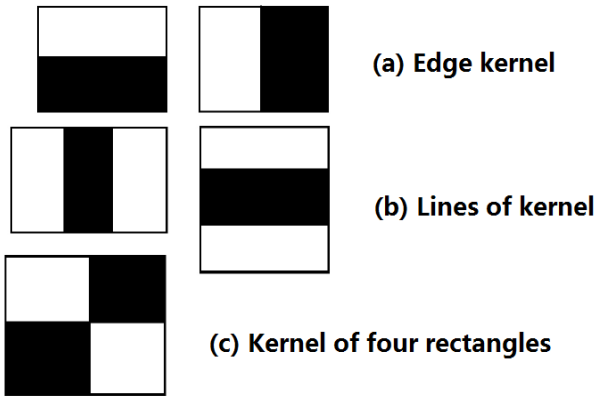

Figure 1. Core for Haar cascade.
Source: Adapted from mordvintsev (2013).

The core is a digital matrix representing the black-and-white part in **Figure 1**. Therefore, during scanning, each pixel in the image is multiplied by the corresponding number of cores. The difference between the white part pixels and the black part pixels produces characteristic values. At the end of projection, a set of features is obtained.

The second stage is based on AdaBoost, which aims to select a small part of visual features with higher intensity for a given image. These visual features are used for target detection and classification, which is based on pulse algorithm. Pulse is a machine learning technology that combines several weak classifiers to improve the overall accuracy[17].

The first selected feature is focused on the eye area, because the eye area is usually darker than the nose and cheek area. The second feature selected is based on the fact that the eyes are darker than the tip of the nose. For each visual feature, this method selects the resource with the lowest error rate to find parameters to classify future positive and negative faces.

The cascade process is carried out in the third stage, that is, the intermediate combination of degenerate tree classifiers. The strategy reduces the false positive rate by combining classifiers at each stage of the cascade. To illustrate this process, an image-based classifier is applied, which discards the images classified as negative, that is, the images without target features. The next step is to consider only the correctly classified images and apply another classifier by rejecting the negative images again. The cascade process continues with an ideal hit rate. **Figure 2**, adapted from Harmouch[18], illustrates the cascade classification process.

Similar to various Supervised learning methods, Haar also needs training stage. To do this, you must create an image library separated by positive (including the image of the object to be recognized) and negative (excluding them). At the end of the training process, the program stores the Haar characteristics of the object in an XML file, which will be used as the model of object detection.

In order to train the classifier, each image gets an equal weight at the beginning. After each classification, the weight of the incorrectly classified image will increase. Then, the same process is performed by calculating the error rate and the new

weight. Repeat this process until the required accuracy is reached or the required number of resources is found according to the error rate.

Haar Cascade is free and located in the OpenCV repository. It used to train classifiers with XML files. Therefore, its integration with the system is simple and flexible, which makes its implementation very popular in artificial vision projects for target detection.

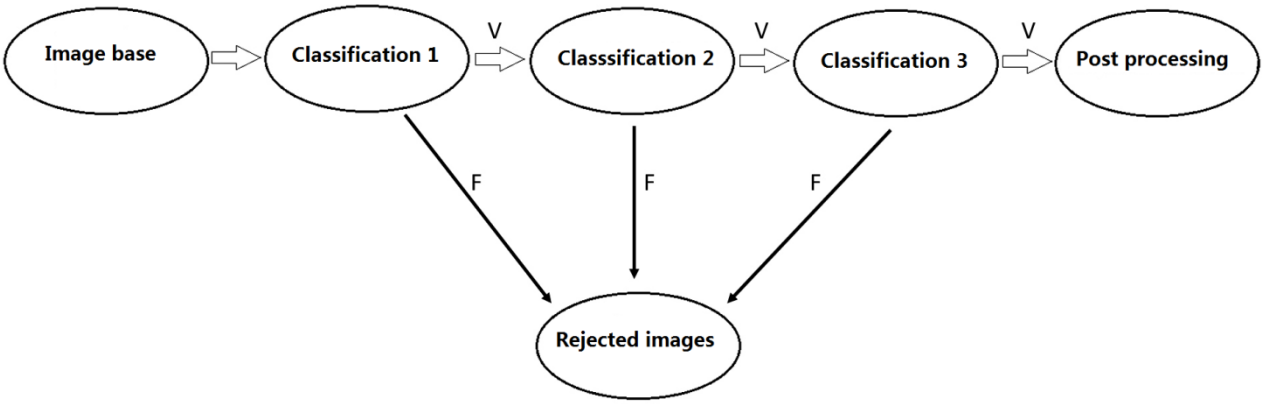

Figure 2. Cascade classification process In each step, a series of images (f) with wrong classification are rejected, and the image (V) with correct classification is then classified next time.
Source: Adapted from Harmouch[18].

3.2 Component diagram

Figure 3 shows the relationship between the application and the OpenCV framework responsible for image processing and coding[19]. Haar Cascade classifier is used to detect patterns in images. The benefits provided by Haar Cascade include ease of implementation and low consumption of computing resources to train and approve models.

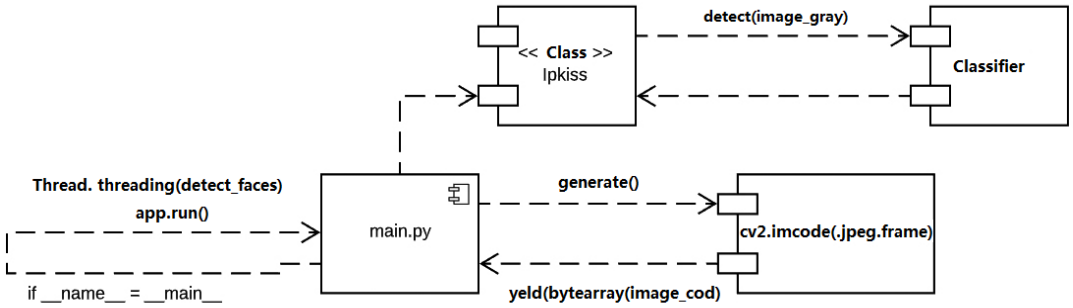

Figure 3. The component responsible for capturing and converting images.
Source: I designed it myself.

Figure 4 shows the connection between the application and MongoDB Atlas database[20]. This is a non-relational database, also known as NoSQL. This type of database improves the speed of transactions and does not give priority to the atomicity, consistency, isolation and persistence (ACID) related attributes in relational databases. MongoDB Atlas is a document oriented database that shares similar functions with other types of banks, such as securities, charts and columns. It also uses cloud computing resources. It is a service characterized by DaaS (database as a service), which allows any computer connected to the Internet to access its database management system and perform operations such as inserting, deleting, changing and querying records. The implementation model allows the flexibility of adjusting the computing resources used according to the needs of the solution, such as the transmission speed and the mode of storing data.

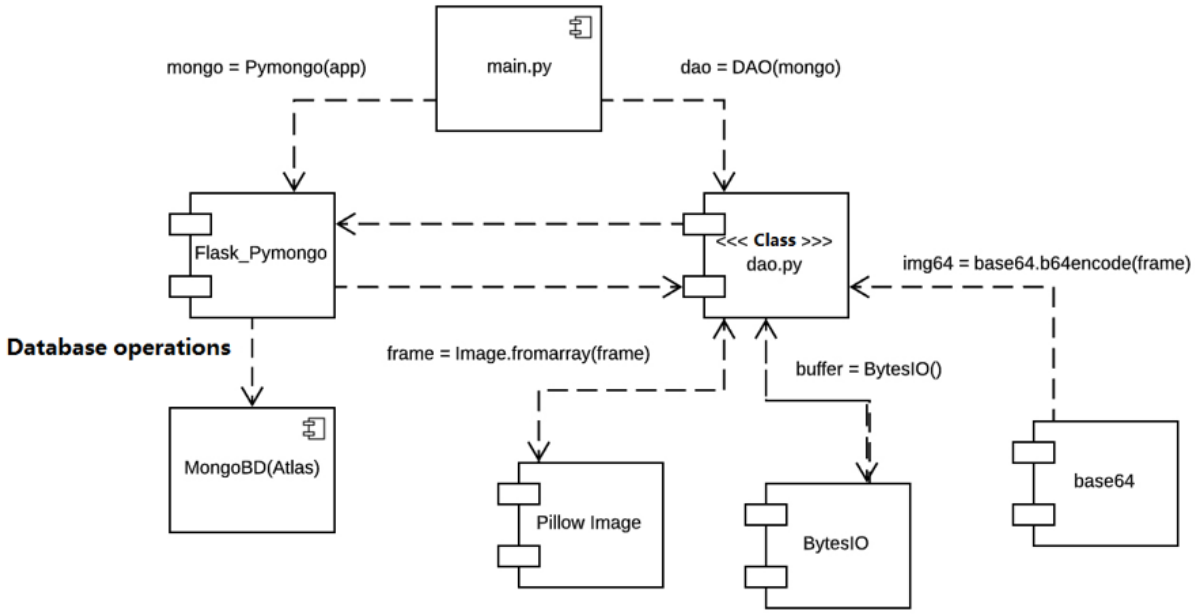

Figure 4. Transaction database component.
Source: I designed it myself.

PyMongo library[21]facilitates the interaction between python programming language and MongoDB database. It contains a set of tools to establish a connection between the display and data layers of a web application. The project uses PyMongo library to build the connection, query, change, insert and delete methods of records in the database.

The images taken by the camera are processed and then entered into the MongoDB Pillow database[22]. Then IO (Python Software Foundation, 2020b) and Base64 (Python Software Foundation, 2020a) were responsible for processing the task.

Figure 5 shows the relationship between the *front-end* application layer components developed in this project. Micro-*framework* Flask[23](Pallets,s.f.) is used to organize applications processed by applications. This *framework* was chosen because it has an optimized structure and sufficient resources to meet the needs of the project, and it adheres to the Python language. Another advantage is that it is modular and flexible, and can serve different types of web applications, from the simplest to the most complex.

Using Javascript libraries improves the usability of applications. Therefore, the jQuery Library[24] is used to create asynchronous licenses, so as to avoid loading pages multiple times during user interaction. Chartist.js library[25]. On the other hand, helps to build responsive and dynamic graphics, while the Materialize Library[26]was created by a group of students at Carnegie Mellon University according to Google's Material Design principles and is responsible for the style sheets of applications.

Algorithm 1 (**Table 1**) provides the functions of capturing, detecting and detecting human faces. The algorithm runs until the user exits the program (lines 2–14); take a camera image (line 3); resize to speed up the classification process (line 4); convert to grayscale (line 5) and create dark and clear areas to identify edge or texture changes; the date and time of the system in which the database record is stored (line 6); face detection using Haar features (line 7). If one or more faces are recognized (line 8), face cutting is performed in the original image (line 10) for each face detected in the image (lines 9 to 12), and the cut faces are stored in the database (line 11).

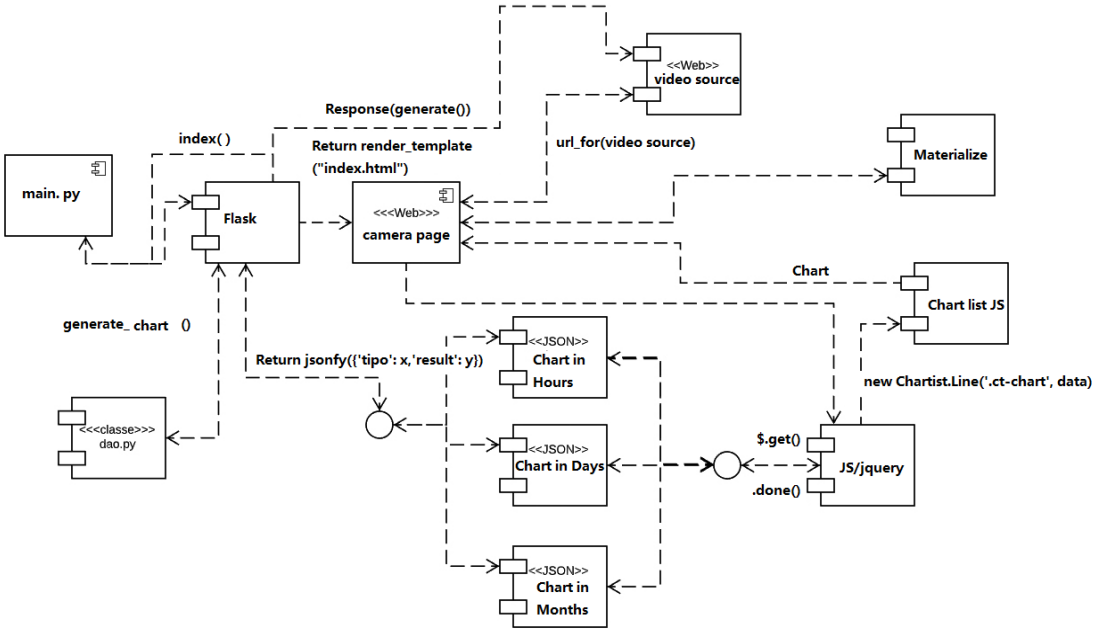

Figure 5. Components responsible for web visualization.
Source: I designed it myself.

Table 1. Algorithm 1 describes the process of detecting and storing faces without masks

Algorithm 1. Function detection and storage

1: **Function** detection
2: **As** long as the user does not complete **the program**
3: Capture new *img*
4: Resize *img*
5: *Convert* image to grayscale *imgg*
6: Store local *date and time*
7: Identify *faces* in *imgg* using Haar Cascade classifier
8: **If** the number of *faces* **>0, then**
9: **For** each *face* ∈ img **do**
10: Cut the face from the *img* and store it in *the face*
11: Stored *in* database
12: **End stop**
13: **End if**
14: **At the end**
15: **end**

Source: I designed it myself.

3.3 Technology used

Table 2 shows a summary of the technologies used in this project. The first column identifies the technology used, and the second column corresponds to the web application layer that requires the technology. When functions are performed on the user's machine, these layers are subdivided into clients, and when functions are performed on the server responsible for receiving requests, these layers are subdivided into servers. The last column provides the reasons for using each technology.

4. Results and discussion

A previously trained Haar Cascade algorithm model is used, which is not available in OpenCV library. The algorithm verification step implemented in the project uses a database containing 194 images (available from DataFlair, 2020). The images used are divided into two categories: one is the face image of people wearing personal protective masks, and the other is the face image of people without masks. Haar Cascade algorithm detects faces with or without protective masks through accuracy, revocation

and f_β-*score* measurement[27].

Table 2. Overview of technologies used

Technology	Layer	Reason
Git	The server	Version control code, save the implementation to. Warehouse. Is required to store items.
Python.	The server	It needs to be integrated with artificial vision and intelligent library.
OpenCV，	The server	Computer vision main library; In all the functions she performs, she is responsible for communicating with cla-actor.
Harr Cascade	The server	A classifier for detecting the presence of faces. A video frame.
Pillow	The server	Data needs to be written to temporary storage.
BytesIO	The server	Temporary storage needs to be created.
Base64	The server	In any server or web application, you need to convert a set of bytes into understandable text.
ImUtils	The server	It needs to capture the camera's IM, agen and cam, biar Tam, ano del m, arc.
Flask	The server	Micro framework for web development.
HTML	Customer	P> Atr6n web page.
Javasctipt	Customer	It needs to execute requests from web applications to servers and databases.
jQuery	Customer	It needs to simplify javascript functions.
Chartist JS	Customer	It needs to create simple, easy to accept graphics.
Materialize	Customer	Responsible for the layout generated by the style sheet.
PyMongo	The server	It is necessary for the application to interact with mongodb.
MongoDB Atlas	The server	Non relational and document oriented database management system.

Source: I designed it myself.

Table 3. Matrix confusion

		Truth value	
		Positive class	**Negative class**
Forecast	Positive class	90 （TP）	54 （FP）
	Negative class	7 (New)	43 (Tennessee)

Source: I designed it myself.

The precision, reversal and f_β-*score* were based on the values obtained by the classifier from the occurrences in measures of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). The appearance belonging to TP category refers to the correct classification number of images with unmasked faces; FP value refers to the case of misclassification in the face image without mask; TN value refers to the incidence of correct classification of face and mask images, and the last FN value refers

to the incidence of incorrect classification of face and mask images. **Table 3** shows the confusion and distribution matrix of these values.

From these values, the accuracy is determined by the calculation of equation 1, where TP is divided by the sum of TP and FP:

$$P = \frac{TP}{TP + FP} \quad (1)$$

Conversely, revocation is defined by dividing the TP evaluation by the sum of TP and FN, as shown in equation 2:

$$R = \frac{TP}{TP + FN} \quad (2)$$

Conversely, the f_β*-score represented* in equation 3[28] is the harmonic average between precision and revocation, and its values represent parameters P and R, respectively. Parameter β can *be* used to assign different weights to the metrics used in the equation. In the performance evaluation used in this project, β value is equal to 1; therefore, accuracy and updating are equally important.

$$f_\beta \, \text{scor}\, e = \frac{\left(1+\beta^2\right)\cdot(P\cdot R)}{\left(\beta^2\cdot P + R\right)} \quad (3)$$

Therefore, through equations 1, 2 and 3, the precision, cancellation and f_1*-score* values shown in **Table 4** can be obtained.

Table 4. Results the accuracy, recall and f_1*score* were obtained

Exact	Revoke	f1-score
0.63	0.93	0.75

Source: I designed it myself.

In addition to evaluating the Haar Cascade algorithm, a minimum viable product (MVP) was constructed using the techniques shown in **Table 1**. The functional prototype of this web application demonstrates the performance of the technology and *hardware* evaluated in this project in a coupled and synchronous manner. **Figure 6** shows the initial screen for developing the prototype. The area marked with blue dotted line displays the video captured by the camera in real time and sends it to the computer vision algorithm. The second area is light green, showing a data carousel with the faces of people without protective masks extracted from the video. The third area, represented by orange dots, shows a graph of the number of individuals who did not wear masks during a specific period (monthly, weekly or daily).[1]

4.1. Discussion

This section discusses the test results, because if copied with other technologies, the results may be different, depending on whether these technologies are appropriate.

First of all, it must be remembered that hyperparameters are sensitive to changes because they control the classification process of images. Therefore, in order to optimize the classifier, you need to explore the parameter space by filtering the parameter space more accurately, which may take hours or even days to find the best super parameter option for your problem.

For example, the detectMultiScale function has two hyperparameters that directly influence the classification: ScaleFactor and MinNeighbours. The first is to determine the size of the image to identify more PE queue objects (in this case, for face detection); this parameter will affect the detection of people far away from the camera, but on the contrary, the false alarm rate and detection speed tend to increase. The second parameter is related to the kernel size (the matrix of adjacent pixels calculated and analyzed), which makes the image clearer and the details more related to the classifier; However, this will require greater processing power.

With regard to image processing, it should be noted that the performance of the classifier may be different by applying other technologies and filters

[1] MVP demo video can be obtained from the following link: https://youtu.be/QmCs_piHZkw. Accessing the solution repository https://github.com/Eskandar1/Ipkiss

provided by OpenCV. For example, it depends on whether the image is balanced or binarized by Gaussian smoothing. These technologies and filters have their own super parameters for optimization.

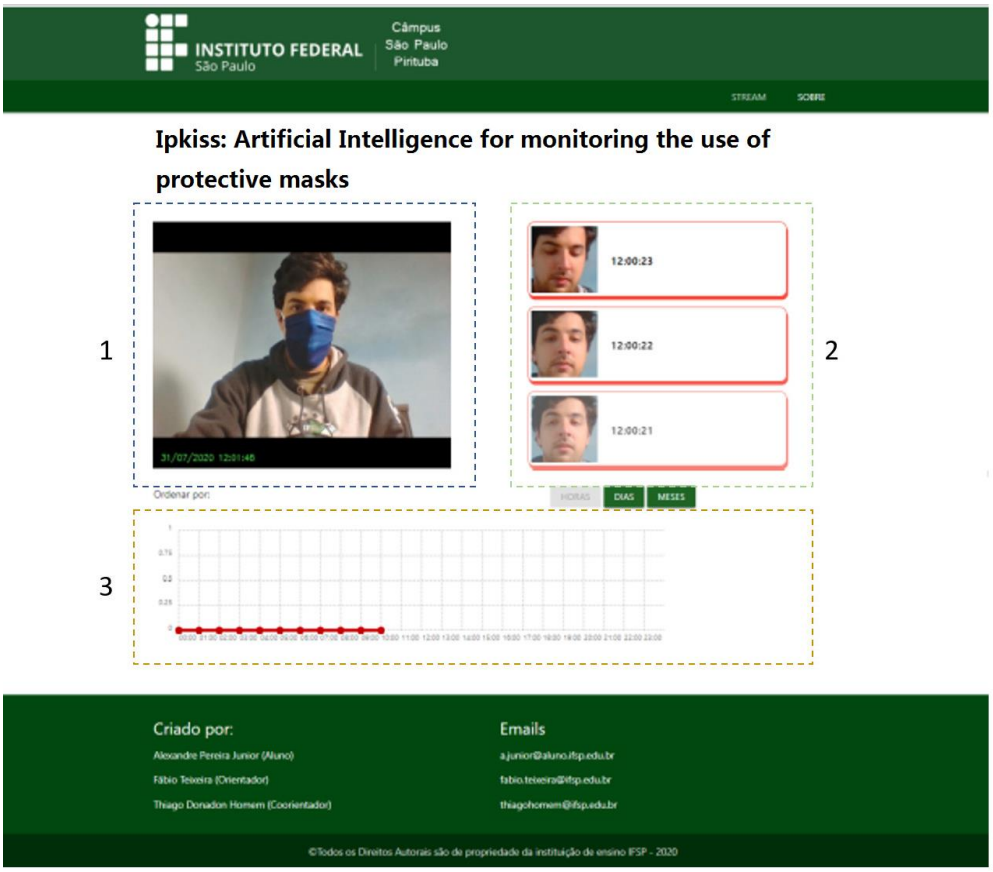

Figure 6. A web prototype system is developed.
Source: I designed it myself.

Another thing worth emphasizing is that the accuracy of f_1-*score* formula is as important as revocation. In addition to accuracy, the algorithm must also detect the maximum number of faces without protective masks in a given scene, which is represented by the f_1-*score* performance measure achieved by the project through the balance between accuracy and revocation. This allows good results to be obtained by MVP analysis of a large number of personnel movements.

One limitation of Haar Cascade classifier is that it can only recognize positive faces. In an uncontrolled environment, people can walk in different directions, so they won't detect faces that don't look at the camera. One way to solve this limitation of the classifier is to use a model with deep learning technology, which contains hierarchical separated neurons, which are specially used for complex pattern recognition in the training process, such as recognizing protective masks, regardless of facial angle.

The data processed during prototype execution is stored in the database. This storage allows them to cross with other sources; for example, public data provided by Datasus, the Information Technology Department of Brazil's single health system (SUS). Such data cross cutting can enhance the analysis of the epidemic, including virus transmission and the use of protective masks.

5. Experiment

Considering the prevalence of COVID-19 and the nursing needs for virus spread, this paper introduces the application of computer vision technology to monitor the use of mask. A web application is developed, which can detect and record the facial images of people without additional wrist guards. This is an application that can be easily implemented on low-cost devices.

The system can detect the face without protective mask with 63% accuracy and record it in the database. This enables those responsible for monitoring the specific environment of implementation to meet the need for effective measures to combat and prevent infectious diseases.

As a suggestion for future work, we intend to use *deep learning* technology to improve the efficiency of the solution. This recommendation must weigh the relationship between facial accuracy and performance so that the application can be implemented on low-cost devices.

Conflict of interest

The authors claim that there is no potential con-flict of interest related to this article.

Acknowledgement

The authors claim that, in order to implement this article, they received financial support from the Institutional Program of Scholarships for Scientific and Technological Initiation of the Federal Institute of Education, Science and Technology of São Paulo.

The authors wish to thank the Federal Institute of Education, Science and Technology of São Paulo (IFSP) for their support in the development of this work. Alexandre Pereira Junior thanks the Institutional Program of Scholarships for Scientific and Technological Initiation of the IFSP for its support in the realization of this article.

References

1. Vieira J, Ricardo O, Hannah C, et al. What do we know about COVID-19? Revista Da Associação Médica Brasileira 2020; 66 (4): 534–540.
2. Reis F. Esteudo avalia efeka Das mascados de prot é no coronaverus. Pfarma; 2020.
3. Hewings-Martin Y. How do SARS and MERS compare with COVID-19? Medical News Today; 2020.
4. Diário Oficial de São Paulo. Decreto 59396 2020 de São Paulo SP; 2020. Available from: https://bit.ly/3hb-hJGbb.
5. Sistema de Monitoramento Inteligente do Governo de São Paulo (SIMI-SP). Isolamento 2020. Governo do Estado de São Paulo. Available from: https://www.saopaulo.sp.gov.br/coronavirus/isolamento/.
6. Pontes R. Inteligência artificial nos investimentos. Clube de Autores (managed); 2011.
7. Backes AR, Junior JJ. Introdução à visão computacional usando Matlab. Alta Books Editora; 2019.
8. Almanza JC, Visão computacional e aprendizagem automática para aplicações em agropecuária e ciências forenses; 2018. Available from: http://www.gpec.ucdb.br/pistori/orientacoes/planos/carlos2018.pdf.
9. Kissler SM, Tedijanto C, Goldstein E, et al. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science 2020; 368(6493): 860–868.
10. Portal do Governo. Governo de SP apresenta Sistema de Monitoramento Inteligente contra coronavírus. Governo do Estado de São Paulo; 2020. Available from: https://bit.ly/2X3ReFn
11. Araujo ZL,Rrosito J C. The deep neural network is used to detect the gender in wild animals in real time. In 2018, the 31st sibrapi graphics, patterns and images Conference (sibrapi). 2018. p. 118-125. https://doi.org/10.1109/sibgrapi.2018.00022.
12. Botelho G, Papa J, Marana A. On learning the deep local features of robust face deception detection. In 2018, the 31st sibrapi graphics, patterns and images Conference (sibrapi). 2018. p. 258-265. Https://doi.org/10.1109/sibgrapi.2018.00040.
13. Finizola JS, Tagino JM, Teodoro FG, Lima CA. (2019). A comparative study of deep face, autoencoder and traditional machine learning technology for biometric face recognition. 2019 International Joint Conference on neural networks (IJCNN). 2019. p. 1–8.
14. Mordvintsev A. Face detection using Haar cascade. OpenCV-Python Tutorials 2013.

15. Viol P, Jones M. Enhanced cascading fast object detection using simple features. Proceedings of the 2001 IEEE Computer Society Conference on computer vision and pattern recognition. CVPR 2001. Https://doi.org/10.1109/CVPR.2001.990517.
16. Lienhart R, Maydt J. An extended set of Haar-like features for rapid object detection. En Proceedings. International Conference on Image Processing 2002. Available from: Https://doi.org/10.1109/ICIP.2002.1038171
17. Duarte JC. O algorithm enhancement at startup and its application. Coleção Digital 2009.
18. Harmouch M. The Haar cascade classifier in opencv intuitively explains this. Medium 2020.
19. OpenCV Team. Opencv 2020. Available from: Https://opencv.org/.
20. MongoDB Inc..Montgobu atlas 2020. Available from: Https://www.mongodb.com/cloud/atlas
21. Dirolf, M. Pimongo: Python driver for mongodb (3.11.0) [software]. Python 2020. Available from: https://pypi.org/Project/pimongo/.
22. Clark A, Lundh F, GitHub contributors. Pillow: Python Imaging Library (7.2.0) [software]. Python 2020. Available from: https://python-pillow.org.
23. Pallets (s. f.). Welcome to Flask's documentation [página web]. Consultado el 14 de agosto de 2020. Available from: https://flask.palletsprojects.com/en/1.1.x/.
24. OpenJS Foundation. Jquery 2020. Available from: Https://jquery.com/.
25. Kunz G. Chart. Chartist.js Simple responsive charts 2019. Available from: Https://gionkunz.github.io/chartist-js/.
26. Chang A, Wang A, Mark A, et al. Documentation [software]. Google 2020. Available from: Https://materializecss.com/.
27. Gehanno JF, Rollin L, Le Jean T, et al. Determine the accuracy and recall of the search strategy of MEDLINE rework study. Journal of Occupational Rehabilitation 2009. 19 (3):223–230.
28. Magdy W, Jones GJ. Chairman: Evaluate the scoring criteria for recall oriented information retrieval applications. Proceedings of the 33rd ACM SIGIR International Conference on research and development of information retrieval. 2010. p. 611–618.

Journal of Autonomous Intelligence (2021) Volume 4 Issue 2.
doi: 10.32629/jai.v4i2.501

Original Research Article

Prediction of Wind Power Generation with Modern Artificial Intelligence Technology

Ibargüengoytia-González Pablo Héctor, Reyes-Ballesteros Alberto, Borunda-Pacheco Mónica, García-López Uriel Alejandro

Instituto Nacional de Electricidad y Energías Limpias, Spain

ABSTRACT

In view of the continuous growth of energy demand and interest in environmental protection, the use of clean energy to replace fossil fuels is a global trend. Wind energy is the fastest growing renewable energy in the world in recent years. However, in the case of Mexico, there are still some difficulties in promoting its use in some areas of the national territory. One difficulty is knowing in advance how much energy can be injected into the grid. This paper introduces the development of artificial intelligence technology for wind power generation prediction based on multi-year meteorological information. In particular, the potential application of Bayesian network in these prediction applications is studied in detail. A weather forecasting method based on Dynamic Bayesian network (RBD) is proposed. The forecasting system was tested using meteorological data from the regional wind energy technology center (CERT) of the National Institute of Electricity and Clean Energy (INEEL) in Oaxaca, Mexico. The results are compared with the time series prediction results. The results show that dynamic Bayesian network is a promising wind power generation prediction tool.

Keywords: *Wind power generation, Power prediction, Artificial intelligence, Dynamic Bayesian network*

ARTICLE INFO

Received: Aug 21, 2021
Accepted: Oct 19, 2021
Available online: Oct 25, 2021

***CORRESPONDING AUTHOR**

Ibargüengoytia-González Pablo Héctor
pibar@ineel.mx;
Reyes-Ballesteros Alberto
areyes@ineel.mx

CITATION

Pablo Héctor IG, Alberto RB, Mónica BP, *et al*. Prediction of Wind Power Generation with Modern Artificial Intelligence Technology and Prediction of Wind Power Generation with Artificial Intelligence Tools 2021; 4(2): 52-62. doi: 10.32629/jai.v4i2.501

1. Introduction

The continuous growth of energy demand and the depletion of fossil and nuclear energy, coupled with the need to protect the environment, make the intensive and extensive use of renewable energy attractive and necessary. Compared with traditional energy, renewable energy has inexhaustible advantages. Its operation cost and the generation of pollution sources, especially the emission of carbon dioxide, are very expensive. However, some of their main disadvantages, such as the intermittent and variability of solar energy, are difficult to predict. These two disadvantages are very important, because in the power grid, in order to meet the requirements of power grid operation specifications, it is necessary to maintain a balance between the power consumption and its power generation. For example, if the electricity generated by wind power shows variability in the amount of electricity generated, non-distributable electricity will be considered, which may lead to quality or power interruption.

Wind power generation is affected by atmospheric parameters, such as wind speed and direction, temperature, humidity and pressure, as well as internal factors, such as maintenance plan and design constraints. On the one hand, it is obvious that the prediction of wind power generation makes the control, management, maintenance and planning of power

dispatching possible. On the other hand, the impact of good dispatching is transformed into significant economic savings and better utiliza-tion of renewable resources by power suppliers. When user demand and supplier scheduling are esti-mated synchronously, the benefit is greater. There-fore, it is best to develop a method to adjust the sup-ply of wind farms according to wind conditions. These forecasts must have such a prediction range to help calculate the power output and improve the fit-ting of the actual demand curve.

Wind power prediction can be realized by phys-ical methods and random methods. All physicists are based on physical considerations of terrain, such as roughness, terrain and obstacles, and the atmos-phere, where they simulate the crazy contours of the wind. In the stochastic model, the prediction is based on the analysis of data series and the use of time se-ries, statistical technology and artificial intelligence technology.

This paper presents an artificial intelligence (AI) method based on Dynamic Bayesian Network (DBN) and machine learning, which is a good wind forecast-ing method. DBN is an extension of static Bayesian network or simple Bayesian Network (BN). Its main features are:

- They allow coding of knowledge and management experience and help users maintain their models to improve their confidence in the accuracy of the models.
- They use multiple model learning algorithms based on historical data to deal with different types of applications.
- They have a powerful reasoning mechanism to respond to information queries given some evidence.
- The output of dynamic Bayesian network is probability distribution, not a point prediction value.

They allow handling noisy or incomplete information and are ideal for intermittent processes.This paper expounds two kinds of original input: academic input and technical input. Academic contribution corresponds to the development of DBN on short-term prognosis. Specifically, it is part of the construction originally proposed by DBN[1]. When calculating data sets in the form of time series, it is extended to prediction applications. Especially in this case, it is used to predict the wind and generate electricity on the 5-hour horizon. This recommendation is demonstrated by predicting the wind speed and direction of Oaxaca, Mexico, using two years of historical data from a wind speed measurement station in the region. On the other hand, technical inputs include the development of proprietary software tools[2], which allow dynamic Bayesian networks to be learned from data. This tool is used to perform the experienments. Finally, the absolute error of the model is used to evaluate the quality of the wind speed. The results are compared with the traditional prediction methods, the average error is acceptable, lower than other methods, and satisfactory results are obtained.

The structure of the rest of this paper is as follows: the next section briefly introduces the prediction of wind power generation. The next section introduces the Bayesian network and dynamic Bayesian network tools for wind power generation prediction. Experiments and results will be described, analyzed and evaluated in the following section. Finally, Finally, the last section concludes the article and presents future work.

2. Wind prediction

The wind power prediction corresponds to the offshore power generation of an aircraft at a certain time in the future. Weather forecasting can be carried out on different time scales. Short term predictions range from milliseconds to a few milliseconds and are used to actively control aerogenerators. Forecasts made within hours or up to 3 days are medium-term and contribute to power system management of the energy system and its trading. These predictions help to determine the use of traditional plants (*unit commitment*) and the optimal scheduling of these plants *(economic dispatch)*. The 5–7 day forecast is called long-term forecast and is used for maintenance planning.

Wind energy is the kinetic energy generated by mass airflow. In this case, it is stable and predictable on an annual scale, but it depends on weather condi-

tions on a smaller time scale. Meteorological conditions produce changes directly in air mass movement according to incident solar radiation, ambient temperature, relative humidity, atmospheric pressure, altitude and latitude. These changes may be cyclical in days, months, seasons, seasons or years. The speed of the wind may be very small, from 2 km/h in the breeze to 120 km/h in the hurricane, lasting from a few seconds to a few days.

Wind energy generates electricity by converting kinetic energy into mechanical energy, which is then converted into electrical energy by wind turbines. According to Bates law, the maximum wind energy available to wind turbines is 59.3%[3], but the maximum wind energy of commercial wind turbines is 75% to 80% of the Bates limit[4]. The available power when the wind passes through a vertical area at a certain speed is determined by

$$p = \frac{1}{2}\rho A v^3 \quad (1)$$

Where ρ is the air density. Within the operating temperature range of Oaxaca, the air density is assumed to be constant. **Figure 1** corresponds to the power curve of the wind turbine and shows the power generation as a function of wind speed. The power curve characterizes the air generator. The minimum speed at which power begins to be generated is called the starting speed and is typically 3 m/s or 4 m/s. With the increase of speed, the power also increases. The wind turbine runs under partial load until it reaches the rated speed and matches the rated power. The wind turbine is designed to produce maximum power at a speed between rated speed and stop speed, which corresponds to the maximum speed at which the wind turbine operates under safe conditions, with a typical speed of 25 m/s.

Wind energy prediction can be carried out directly or indirectly. In the first case, the estimation is realized by directly describing the power output of the electrical power variables. In the second case, it predicts the behavior of wind by estimating meteorological variables and the correlation between power curve and electric power.

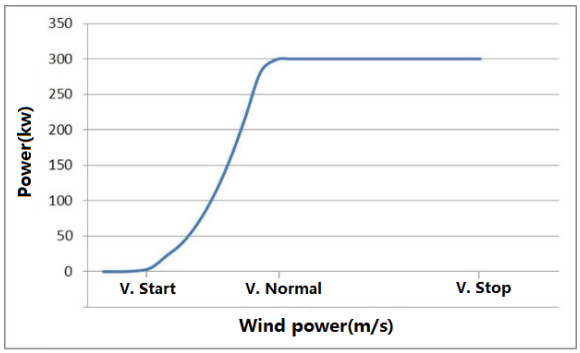

Figure 1. Wind turbine power curve

The models used for prediction can be divided into two categories[5]: physical models[6]and stochastic models[7]. Physical methods are based on physical considerations of terrain, such as roughness, terrain and obstacles, and atmosphere, and simulate local wind profiles. These models include numerical weather prediction (NWP)[6], sky image analysis[8]and power generation system

In the stochastic model, the prediction is based on data sequence analysis and carried out through various technologies:

Time series. When data is available, they use historical data. The prediction of variables is completed by multiple transfer values of the same variable.

Statisticians. They use statistical functions to estimate the value of a given variable and the historical data of that variable and other related variables.

Artificial intelligence (AI). This includes building models using automatic learning algorithms, expert knowledge, or a mixture of both.

The main statistical methods include AR, ARMA and ARIMA[9]. On the other hand, the current wind field prediction models based on time series and statistical models include ALEASOFT, AEOLIS, CASANDRA, CENER, GARRAD HASSAN, METEOROLOGICA, METEOSIM or METEOTEMP[10]. They are based on the prediction of atmospheric changes by some numerical models, and do not have enough accuracy to predict the horizon wind speed of more than 5 hours. As another option to solve this problem, we can find a method to solve complex problems that cannot be solved by traditional methods by using artificial intelligence tools

such as artificial neural network (ANN)[11], Bayesian network[12], fuzzy logic and support vector machine[13]. These methods "learn" the relationship between prediction and measurement series. These methods usually provide better results within 2–4 hours, depending on the method selected. In addition, artificial intelligence methods usually provide better results than statistical methods.

An effective prediction technology developed by artificial intelligence community is Dynamic Probabilistic Graphical Models (DPGM). Many of the problems listed can be solved by using this method. Its main features are:

(1) Represents a conditional dependency between variables.
(2) This is a kind of reasoning technology with uncertain process and environment.
(3) It can represent the structure and parameters of expert knowledge.
(4) It represents the result in the form of probability distribution, not in the form of point value.
(5) Because MGPDs has the ability to express conditional independence, it implicitly excludes irrelevant variables.

Despite these characteristics, in order to support the assumption that this technology can achieve good results, different models must be established and evaluated. The main DPGMs are Bayesian networks and Markov networks.

In this paper, Bayesian Dynamic Network (BDN) is used to predict wind speed and direction, as described below. BDN links wind speed and direction with ambient temperature, relative humidity and solar radiation to predict the probability of wind speed and direction in the future. Then, the power generated by the power characteristic curve of the wind turbine is mapped. Then it briefly introduces the method of predicting wind power generation.

3. Bayesian networks and Dynamic Bayesian Networks

The problem of wind power generation is the variability of its power fluctuation and availability. I mean, it's an uncertain question. Among the ideal artificial intelligence methods for dealing with uncertain problems, Bayesian network (BN) has been proved to be practical in general practical applications, especially in alternative energy[14]. BN represents the dependent and independent relationship between process or system variables. They are based on Bayes' theorem, which links the conditional probability of events or assumptions H of given evidence E and P (H|E) *with* the conditional probability h and P (H|E) of given evidence E,

$$P(H|\ E) = \frac{P(E|\ H)P(H)}{P(E)} \qquad (2)$$

BNs are AI methods because they allow the knowledge of certain processes to be represented in two ways. First, the network structure represents the dependence and independence between variables.Second, the parameters represent the quantitative knowledge of the process. Parameters refer to the terms $P(E|H)$ and P (H). The term $P(E|h)$ is easily found in historical data, such as disease and symptoms, equipment failure and impact measurement. Then $P(H/E)$ is given by equation (2).

Formally speaking, BN is a directed acyclic $G = (N, E)$, that allows the representation of knowledge in applications dealing with uncertainty[12,15]. A N node represents a set of $X = \{X_1, X_2, \ldots, X_n\}$ random variables. The arcs E in the structure represents the probabilistic relationship between nodes. If $Pa(X_i)$ is the parent node set X_{i^*} of a node, the Bayesian network structure corresponds to the joint probability distribution of the application, as shown below:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i|\ Pa(X_i)\right) \qquad (3)$$

Among $Pa(X_i)$ them, it represents the parent node of the variable node, which further means that for each variable in X_i the network, BN, RB, X_i when it has the information of the parent node, it is conditionally independent of the non child nodes in $\mathrm{Pa}(\mathrm{X_i})$ the network. In summary, BN represents the dependency and independence between variables in the application.

For dependencies, they are quantified in the form of the Conditional Probability Table (CPT) of the child node value, given the value of the parent node $P(E|H)$. All TPCS and a priori probability vectors P (H) of child

nodes need to be provided.

Once the process knowledge is captured in the analysis, the probabilistic inference process can be performed, that is, the evidence *e* is used to assign a value to the cone node, and the probability of some assumptions *P* (*H*|E = e) can be calculated. This is equivalent to calculating the marginal probability of the unknown variable, given the known variable *P* (*X*|*e*)).

The BN described so far infers according to the information at a given time point and the evidence in the model. In other words, there is no time dependence. However, some applications, such as wind forecasting, want to establish conditional models between current and past meteorological variables to calculate future values. In order to consider weather factors in Bayesian prediction model, dynamic Bayesian network (RBD) is developed.

In dynamic applications, the working universe is not only a variable, but also a time series variable that changes with time. Space now consists of $X^{(t)} = \left\{X_1^{(t)}, X_2^{(t)}, \ldots, X_n^{(t)}\right\}$ a $X_i^{(t)}$ collection of X_i time t variables. Then, the prediction $P\left(X_i^T \mid X_j^{(t)}\right)$ problem X_i^T becomes a time $T, i \neq j$ variable.

Obviously, the spatial combination of variables at different time points represents a very complex set. Therefore, the assumptions are as follows:

(1) They are considered time intervals, or discrete [1] time. Therefore, $X^t = X^{(0)}, X^{(1)}, \ldots, X^{(t)} y$ therefore. $P\left(X^{(0:T)}\right) = \prod_{t=0}^{T-1} P\left(X^{(t+1)} \mid X^{(0:t)}\right)$ This means that some values of the T future depend on all values of the past $(0:T)$ and present.

(2) The system is Markovian. This $P\left(X^{(0)}, X^{(1)}, \ldots, X^{(\mathrm{T})}\right) = . \prod_{t=0}^{T-1} P\left(\mathbf{X}^{(t+1)} \mid \mathbf{X}^{(t)}\right)$. In other words, the future has nothing to do with the past and the present.

(3) The system is stationary. That's $P\left(X^{(t+1)} \mid X^{(t)}\right)$ it. Everything's t is the same. This means that the state of the next process depends on the current state, just like in any part of the sample (every year, spring or winter).

The BDN creation program developed in this project was inspired by the BDN proposal[1]. This mechanism, as well as the assumption that the system is Markovian, allows the creation of a two-stage Bayesian Network $P\left(\mathbf{X}^{(t+1)} \mid \mathbf{X}^{(t)}\right)$ to define t anything in the process. This network is called a transition network. **Figure 2** shows the network transformation learned in the case study. Each node represents the meteorological variables involved in the analysis, in which the left layer corresponds to the weather and t the right layer corresponds $t+1$ to the weather. Note that all variables depend on the previous time, and some variables, such as temperature (Temp_1), depend on the current value of solar radiation in addition to the previous Relative Humidity (RH) and Solar Radiation (SR).

Under the third assumption, the system is such that the BND required for future unit time prediction is realized by expanding the transition network at (N + 1) layer or stage. For example, if time data is used and needs to be predicted to 5 hours, BDN is formed by expanding the transmission network to layer 6. Considering that the first layer corresponds to the input layer and each of the five layers contains a prediction for the next hour, it is considered to be six layers. **Figure 3** shows the BDN results of wind forecast on the 5-hour horizon.

To sum up, the learning process of 5-hour power prediction in this project is summarized as follows:

(1) Divide the historical database into two groups. Data for training NBD and data for verifying NBD performance. Note that the data must be a time series, not just a set of records.

(2) If the time base of the original data is less than the required time base, the average value is mapped to the variable value in the required step.

(3) Use uniform division to discrete continuous values.

[1] The symbols used are as follows. Bold capital letters represent variable sets, such as ***X.*** capital letters represent variables, such as X.

(4) Copy training data in two columns of NAS, including records in each row *(x* (T), *X* (T + 1)). The learning algorithm of Bayesian network is used to construct the transition network. In this project, the hugin package[16] was used to build the network in **Figure 2.**

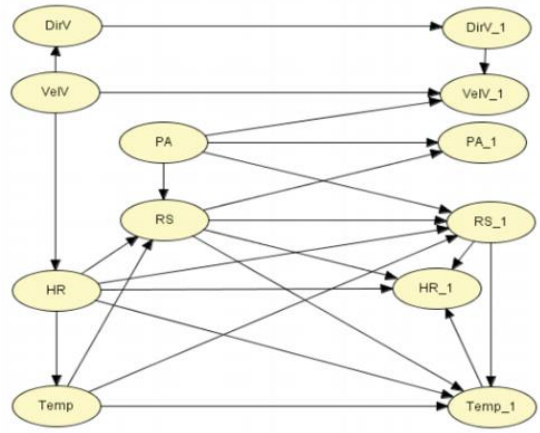

Figure 2. Experimental case study of network transition

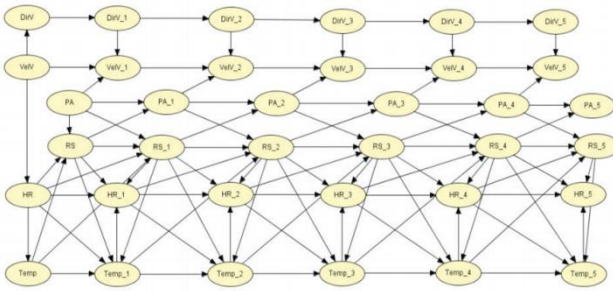

Figure 3. Dynamic Bayesian network results for 5-hour prediction

(5) Expand the transition network on the (N + 1) time slice to form the BND as shown in **Figure 3**. As mentioned earlier, for the horizon with N stages in the future, the (N + 1) layer is required.

(6) The EM algorithm (Expectation-Maximization)[17] is used to learn the model parameters corresponding to the child node conditional probability matrix and the root node prior probability vector in **Figure 3**.

The size of BND increases in layers according to the stages of the predicted horizon. However, this method is limited to the prediction of small stages in the future, because more than 10 or 12 stages will make the Bayesian model impractical due to the number of nodes in reasoning[18].

The reasoning in the model includes assigning values in the variable layer of time *t* and predicting the probability of the sixth layer in the network. The result is the posterior probability distribution of all variables including wind speed.

4. Experiments and result

These experiments were conducted at INEEL regional wind energy technology center (CERT) in Ventosa, Oaxaca, Mexico. Its infrastructure is designed to install up to 5 MW of wind power, which can be integrated with different capacities and models of wind turbines.

CERT sells electricity produced by Japan's 300 kW KOMAI wind turbine, which was donated to INEEL by the global environment facility (GEF) through the United Nations Development Programme (UNDP).

The historical wind data and other meteorological variables obtained from the center consist of time series marked with date and time, which lasted for more than two years. The information collected includes: Ambient temperature (Temp), Relative Humidity (RH), Solar Radiation (SR), Wind Direction (DirV) and wind speed (VelV) at two different heights on the ground. Record the data every ten minutes. The data of 2012 and 2013 are used for training, and the data of January and February 2014 are used to test our system. Preliminary results are presented in[19].

As mentioned above, the model is learned from the weather data of CERT from January 2012 to December 2013. The model is shown in **Figure 3**. The experiment was tested on the time data from January to February 2014. In order to evaluate the performance of the prediction system according to specific weather conditions, experiments under charging current conditions were carried out in a specific time. For example, we loaded the evidence at 0:00 (midnight) and compared it with the prognosis at 5:00. The experiment was conducted from 12:00 to 17:00 (half a day). There are significant differences in solar radiation and temperature between the two periods.

The experiment was carried out by the following methods:

(1) Obtain historical data from the meteorological record variables of a location, mainly including speed and wind direction.

(2) The learning prediction model follows the method described in the previous section.

(3) Use the historical test data of to learn BND feedback with value at a given time point. Through probability propagation, the future posterior probability distribution of N layer wind speed is obtained. This is achieved through the *Hugin* package or any other probabilistic model processing package.

(5) Numerical calculation and comparison with the measured data file. This will be described in detail below.

(6) Difference calculation and prediction error estimation.

Alternatively, RBD training (step 3) can use the current value to predict the next N hours (or stages) in some cases. It is worth mentioning that the prediction error is calculated after n stages are completed.

Figure 4 shows the wind speed forecast results in February. The vertical axis represents the wind speed (M/s). The horizontal axis represents an example of the experiment, once every hour every day. The lines with "Measured" and "Estimated" represent the measured and predicted wind speed do, respectively. As shown in the figure, the predicted value is very close to the actual value in some times, while the difference is more significant in others.

The deviation between the predicted value and the measured value is quantified by error. The inertial measurement error is calculated by the following formula:

$$E_{\text{Inst}} = \frac{\left(VelV_{\text{real}} - VelV_{\text{foreast}}\right)}{\left(V_{\max} - V_{\min}\right)} \times 100 \qquad (4)$$

Velocity $VelV_{\text{real}}$ measurement and VelV prediction are predictive. Instrument engineers use this mechanism when evaluating equipment. This value is divided by the difference between the actual value and the predicted value by the full scale of the instrument. In this project, we assume $V_{\max} = 25\text{m/s}, \text{y}V_{\min} = 0\text{m/s}$.

Wind speed is a difficult variable to predict because of its uncertainty and natural volatility. The average error of 0–5 hours is 8.21%, and the average error of 12–17 hours is 5.76%. In **Figure 4**, the maximum error is more than 19%, while the minimum error is almost 0. Most errors are less than 5%. Although this seems to be a huge error in somecase, the literature shows that a wind prediction system with this average error is promising[20].

The literature suggests that in addition to the measurement of single point prediction error, the uncertainty of probability prediction should also be considered[5,21]. These references suggest the necessity of calculating probabilistic models given numerical weather prediction (NWP) models. Because our prediction mechanism is probabilistic, we get the probability distribution of wind speed under given weather conditions in the next 5 hours. To calculate the NWP, the expected value of the posterior probability vector was used, that is, $V_{\text{est}}=s_n\ VelV_iP_i$ where $VelV_i$ is the central value of the interval i, and P_i is the probability of the interval.

The uncertainty estimation of wind forecast is calculated by Quantile mechanism[20]. The posterior probability distribution of a given wind speed, $P_{t+k'}$ where t the current time is k the current time and is the prediction range (the number of previous q_{t+k}^{α} time slices), $\alpha \in [0,1]$ and the quantile with parameters x is $\text{prob}\left(P_{t+k} < x\right) = \alpha$ defined as the given value.

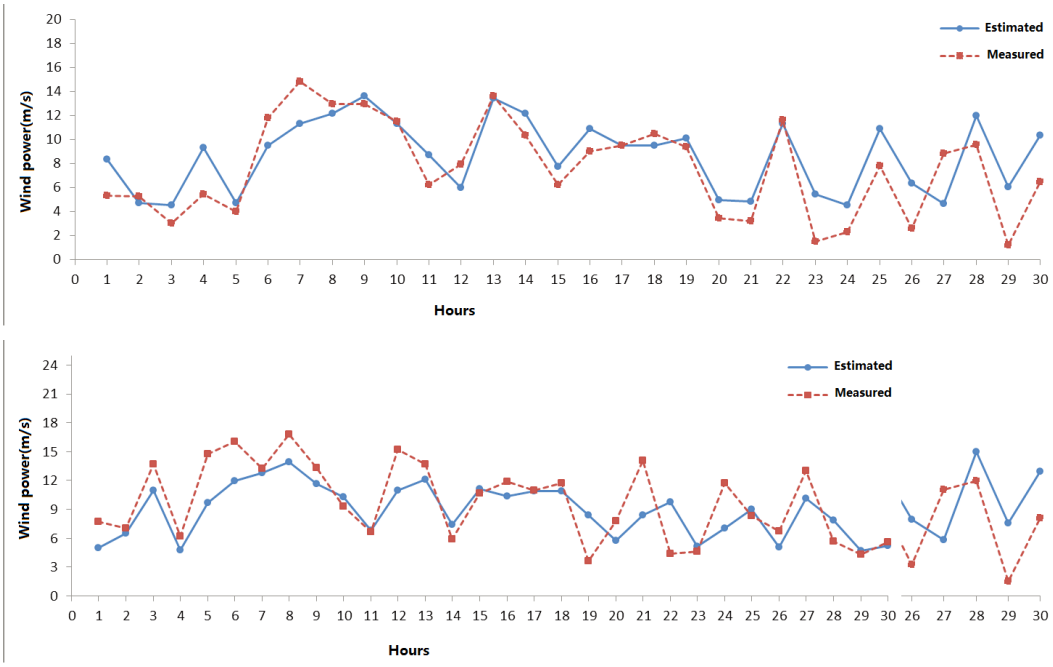

Figure 4. The experimental results of wind speed in 0-5 hours and 12-17 hours are predicted by partial plot. The lines marked "Measured" and "Estimated" represent actual and predicted values. Actual and predicted values are in M/s.

Figure 5 shows the results of the same experiment as **Figure 4** with the two quantiles of 20% and 80%, forming a confidence interval with a probability of 60%. The mechanism stipulates that the expected power generation of a given horizon is 1 to 1.6 MW, with a probability of 60%. In order to complete the evaluation of the experiment, **Table 1** shows the different error measurements of the above two experiments[21]. As we have observed, the experiments carried out from 12 to 17 show small errors. On the other hand, another measure for calculating the error in the prediction problem is the average quadratic error (RMS), which is consistent with the average error of prediction, which is defined as (Osman, 2001):

$$rms = \left[\frac{1}{n} \sum_{i=1}^{n} E_{\text{lnst}}^2 \right]^{\frac{1}{2}} \quad (5)$$

Where E_{lnst} is the instrument error defined in equation 3.

The errors reported in **Table 1** are derived from the data of 30 cases of 0 hours (or 12 hours), i.e. 1 month. In each experiment, the actual value was compared with the predicted value. Calculate the error using equations 3 and 5. A negative number in the first line indicates overestimation because the estimated value is greater than the actual value. **Table 1** compares the predicted values of different indicators of the two experiments. Efforts during the day are better than efforts at night.

In order to compare our prediction results with other traditional time series statistical methods, **Table 2** shows the results of these methods, which are obtained from a set of experience in the same scene and data set.

Table 1. Experimental result error

Error type	Experiment 0-5h	Experiment 12–17 hours
Percentage	-4.4	-2.3
Absolute percentage	8.21	5.67
Minimum value	0.17	0.18
Maximum	19.4	17.36
Mean square error	9.84	7.28

Table 2. Experimental result error of traditional method

	AR model command	Ma command	Map	Reach
Human Resources Department	44	0	0.235171	45
Arms	12	30	0.271023	52.5
Alima (a)	6	27	0.365057	27.5
Alima (b)	6	27	0.280259	35

(a) Average adjustment
(b) Do not adjust the average

The first column shows the experimental model. AR is a 44 order autoregressive model. ARMA is an autoregressive model of moving average (MA), with AR of order 12 and MA of order 30. The latter two models are the 6th and 27th order AR integrals of AR and MA, respectively[22]. The difference between the two methods is that in data preprocessing, the average value is subtracted from all values to predict time invariance (adjusted to the average value). The fourth and fifth columns are absolute error - half percentage point (MAPE) and directional accuracy (DA). MAPE and DA measure the accuracy of the model and predict the future accuracy respectively. Both indicators are expressed as percentages. Ideally, a good prediction model can obtain low map and high DA. MAPE and DA are calculated by comparing the predicted value with the actual value. **Figure 6** shows the results of a short experiment using time series statistics.

By qualitatively comparing the BND results (**Figure 4**) with the results of the time series method (**Figure 6**), it can be seen that the fact of using multiple variables rather than historical variables in the prediction represents a performance advantage. **Figure** 4 shows that the curve tracking between measurement and prediction is larger than that in **Figure 6**.

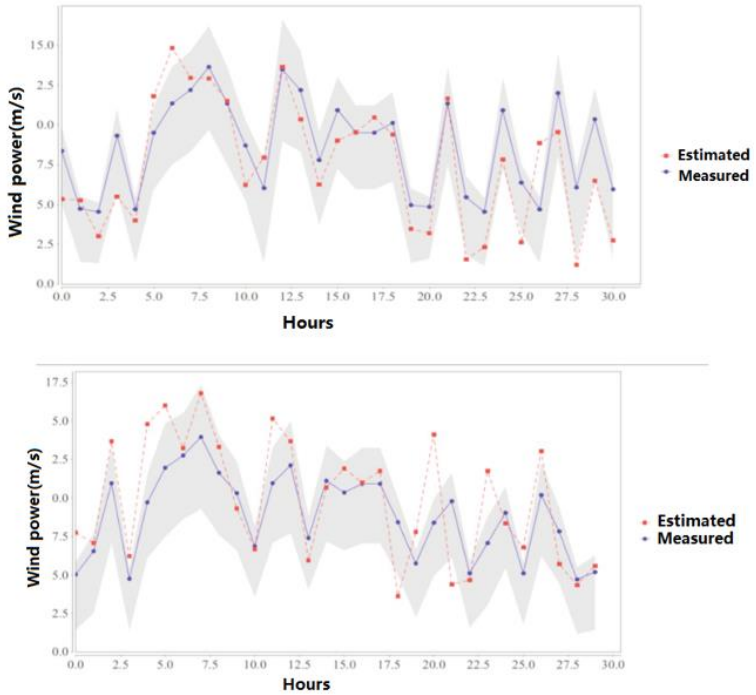

Figure 5. The experimental results are 0–5 hours and 12–17 hours, with certain uncertainty. The area shown represents a 60% probability. Red indicates the measured speed and blue indicates the estimated speed.

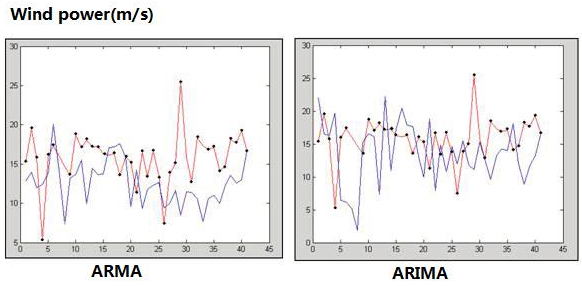

Figure 6. ARMA and ARIMA methods are used to match the experimental results. The red line represents the actual value and the blue line represents the predicted value. Horizontal axis timing.

5. Conclusions and future work

The prediction of wind power generation is an inevitable requirement for the expansion of clean power generation. Bayesian network is a technology used to deal with intelligent systems with uncertainty. This paper presents a new application of dynamic Bayesian network in wind power generation prediction. The most important contribution of this paper is the development of dynamic models dedicated to short-term power forecasting and the methods of learning these models. The innovation of our BND lies in the corresponding relationship between the assumption of prediction problem and the formation of BND structure, which corresponds to the BND classical structure proposed by Murphy[1].

A method of building wind prediction model using dynamic Bayesian network and multi-layer perceptron is proposed[23]. The two prediction models are evaluated and compared to select the model with the highest performance. In the case of Bayesian model, it is recalibrated recursively to restore the accuracy lost due to variable discretization. The novelty of this method is that it compares two techniques with different property variables (discrete variables and continuous variables) and different strength. In this way, it can make better use of the power of approximate functions and methods to deal with uncertainty.

This paper introduces the development of the mathematical formula of dynamic Bayesian network. Preliminary results were published in Ibargüengoytia [19]. In addition, the construction method and theory of Bayesian network proposed by us are described in detail. This paper also describes the measurement error, and discusses the supplementary use of Bayesian model and artificial neural network.

The experiment is carried out on the data of CERT and INEEL meteorological station in La Ventosa, Oaxaca. The results show that artificial intelligence is very helpful in solving the problem of renewable energy. Artificial intelligence provides learning and knowledge representation mechanism, which can be transformed into more effective problem solving methods.

This research work is a preface to a broad theme. Future work in this area will focus on analyzing whether there is any place nearby that can provide useful information for improving prediction, that is, developing dynamic spatial models to improve prediction performance. In addition, the condition that the prediction process is a Markovian process will be studied, and the consequences of this restriction will be analyzed. Finally, the time series will be reviewed to determine whether they are time series, and if not, a BND model will be defined for each stage. This method is compared with other nonlinear time series prediction models.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgement

This project was partially supported by CEMIE-Eolic projects 174919 and P12, both from the SENER-Conacyt Sustainability Fund, and the National Institute of Electricity and Clean Energies (INEEL). Mónica Borunda thanks Conacyt for her position as chair researcher number 71557, and INEEL for her hospitality.

References

1. Murphy KP. Dynamic Bayesian networks: representation, citation and learning [PhD thesis]. Berkeley: University of California; 2002.
2. Ibarguengoytia PH, Reyes A, Romero Leon I, *et al.*

A tool for learning dynamic Bayesian networks for forecasting. Artificial intelligence and its application. Lecture notes in Computer Science 2015; 9414: 520–530.

3. Betz AN. In : Introduction to flow machinery theory. Oxford: Pegamon press; 1996.
4. Zhang Y, Wang J, Wang X. Overview of probability prediction of wind power generation. Renew. Sustain. Energy Rev. 2014; 32: 255–270.
5. Olaofe ZO. A surface-layer wind speed correction: A case-study of Darling station. Review Energy 2016; 93: 228–244.
6. Sideratos G, Hatziargyriou ND. An advanced statistical method for wind power forecasting. IEEE Transactions on Power Systems 2007; 22(1): 258–265.
7. Chow CW, Urquhart B, Lave M, *et al*. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. Solar Energy 2011; 85(11): 2881–2893.
8. Enke W, Spekat A. Through classification and regression, the output of climate model is extended downward to local and regional climate elements. Climate Research 1997; 8: 195–207.
9. Pearl J. Probabilistic reasoning in Intelligent Systems: plausible reasoning networks. San Francisco, DA: Morgan Kaufman Publisher; 1988.
10. Vapnik VN. Statistical learning theory. Willie science.
11. Borunda M, Jaramillo OA, Reyes A, *et al*. Bayesian networks in renewable energy systems: Double blind survey. Renewable and Sustainable Energy Reviews, 62: 32–45.
12. Sukar LE. Probabilistic graphical model: Principles and applications. London: Springer villag; 2015.
13. Andersen SK, Allison KG, Jensen FV, *et al*. Hu Jin is a shell that builds a Bayesian belief world for expert systems. In: Proceedings of the 11th International Joint Conference on artificial intelligence. 1989. p. 1080–1085.
14. Lauritzen SL. EM algorithm of graphical association model without data. Computational Statistics & Data Analysis 1995; 19: 191–201.
15. Cooper GF. The computational complexity of Bayesian network probabilistic reasoning. Artificial Intelligence 1990; 42: 393–405.
16. Ibargüengoytia PH, Reyes A, Romero Leon I. *et al*. Wind power forecasting using dynamic Bayesian models. Advances in soft computing, Part II-MICAI 2014. Lecture notes in Artificial Intelligence, Springer Velag; 2014. 8857: 184–197.
17. Monteiro C, Bessa R, Miranda V, *et al*. Wind power forecasting: State-of-the-art-2009. Portugal.
18. Bremnes JB. Probabilistic wind power forecasts using local quantile regression. Wind Energy 2004; 7(1): 47–54.
19. Osman EA, Abdel-Wahhab OA, Al-Marhoun MA. Prediction of oil PVT properties using neural networks. Society of Petroleum Engineers 2001.
20. Box GEP, Jenkins GM, Reinel GC. Time series analysis, prediction and control. Prentice Hall. Eng-Lewood Cliff.
21. Reyes A, Ibarguengoytia PH, Jijón JD, *et al*. Wind power forecasting for the Villonaco wind farm using AI techniques. In: Lecture notes in computer science. Springer; 2016.

Journal of Autonomous Intelligence (2021) Volume 4 Issue 2.
doi: 10.32629/jai.v4i1.502

Original Research Article

Comparison and Selection of Artificial Intelligence Technology in Predicting Milk Yield

Rudibel Perdigón Llanes[1]*, Neilys González Benítez[2]

[1] *Empresa Comercializadora "Frutas Selectas", Km 1 ½ Carretera a San Juan y Martínez, Pinar* del Río, CP 20100, *Cuba*

[2] *Centro Meteorológico Provincial CITMA, Colón # 106 entre Antonio Maceo y Virtudes, Pinar del Río, CP 20100, Cuba*

ABSTRACT

Forecasts are an effective decision-making tool, mainly in the dairy industry, because they help improve herd management, save farm energy and optimize long-term capital investment. The application of artificial intelligence technology to predict milk yield is a subject of concern in the scientific community. However, defining a technology or model to predict the effective performance of these products in different environments is a challenging and complex activity, because none of them is accurate in all scenarios. This study compared the application of artificial intelligence technology in milk yield prediction in the literature, and applied analytic hierarchy process to select the most suitable artificial intelligence technology for milk yield prediction. Methods comprehensive analysis, investigation and experiment were used. The results show that the artificial intelligence technology based on artificial neural network is more suitable for the prediction of milk yield than decision tree and support vector machine. In the field of milk production, the most relevant selection criteria are identified as the ability of these technologies to process uncertain data and their ability to obtain accurate results in the best way. The analysis carried out supports the decision-making of milk production organization.

Keywords: *Multi-criteria analysis; Analytic hierarchy process; Prognosis; Policy decision*

ARTICLE INFO

Received: Sep 11, 2021
Accepted: Nov 13, 2021
Available online: Nov 19, 2021

***CORRESPONDING AUTHOR**

Rudibel Perdigón Llanes
rperdigon90@gmail.com;

CITATION

Perdigón Llanes R, González Benítez N. Comparison and Selection of Artificial Intelligence Technology in Predicting Milk Yield 2021; 4(2): 63-71. doi: 10.32629/jai.v4i2.502

1. Introduction

Decisionmaking is an analytical process in which the best scheme is selected from multiple schemes in order to achieve a specific goal[1]. This activity is widely used in organizations because managers often need to make decisions on different issues[1,2]. The decision-making of milk production organization is a key factor to improve its production and economic indicators[3]. However, this process is sometimes carried out in the absence of information, which needs to improve its efficiency and accuracy[3].

Forecasting is a useful tool for decision-making in the dairy industry[4,5]. Accurate prediction of milk production can help improve the financial planning of milk producers and avoid economic losses[6-8]. In addition, it can improve cattle management, save farm energy and optimize long-term capital investment[4,5].

At present, different models are used to predict milk production[9]. These models are based on the application of mathematical, random time series, regression and computational algorithms based on artificial intelligence (AI) technology[9,10].

The use of AI technology in agricultural and livestock activities is a growing phenomenon that helps to improve data processing and the profitability of the sector's business[10-12]. Artificial intelligence techniques used in the literature to predict milk production include:

(1) Artificial Neural Network based technology (ANN)[5,10,13-19]

(2) Technology based on Support Vector Machine (SVM)[4,8,20]

(3) Genetic Algorithm based technology (GA)[4,19,21]

(4) Decision Tree based technology (DT)[8,22-24]

These techniques exhibit appropriate behavior for prediction in different fields because they have no restrictions on processing a large number of data and input variables, and can identify, learn and approximate data characteristics by simulating the inherent and nonlinear relationships existing in the data[11,25,26]. In addition, they can obtain accurate results by saving time and computing resources[8].

Artificial intelligence technology enables the agricultural sector to solve specific problems and make effective predictions in an environment with specific characteristics and variables based on its analysis[28]. However, milk production is influenced by various factors that hinder the prediction of milk production[4,9]. Climate change, livestock management, feeding, genetic and physiological characteristics and the incidence of diseases are some factors affecting milk production and complex prognosis[9]. Defining a model or technology to predict the effective performance of these products in different environments is a challenging and complex activity, because accurate models or technologies are not required in all scenarios[5,17,28].

Determining the AI technology most suitable for milk yield prediction is a concern of the international scientific community[8,25,29]. However, in the literature, there is no consensus on the best performance of artificial intelligence technology in this process. In addition, the identified authors take their prediction accuracy as the basic standard for evaluating the effectiveness of these technologies. This factor depends on the environmental factors (prediction variables) for prediction, which may be insufficient according to the characteristics of the decision-making environment (time, available data and resources, interpretability and applicability). The diversity of algorithms applicable to each AI Based Prediction Technology identified in the literature complicates the use of statistical analysis as a method to determine the technology most suitable for these product predictions.

Multi criteria decision making (MCDM) has proved satisfactory results in different application fields, especially when decisions need to be made between different alternatives according to different selection criteria or views[30]. Analytic hierarchy process (AHP) is one of the most commonly used MCDM methods in the literature because of its relevance and practical application[31,32], which is mainly used in Agricultural Research on the selection of production mode of dairy farms[33]. Its main advantage lies in dealing with quantitative (information and data obtained) and qualitative (views of decision makers and characteristics of decision environment) in the decision-making process[31,32,34].

The purpose of this study is to compare the application of artificial intelligence technology in milk yield prediction, and select the most suitable method for milk yield prediction by using analytic hierarchy process according to different selection criteria.

1.1. Selective analytic hierarchy process

It provides effective mathematical support for the analysis of selection problems; it measures quantitative and qualitative standards through a common scale; it allows errors in the evaluation process to be verified based on the inconsistency index and allows the results to be supplemented by other mathematical optimization methods[35]. This study is divided into three basic parts: materials and methods or calculation methods. The part describes the scientific methods and methods used in the research; The results and discussion section describes the main findings obtained in the application of AHP and its comparison with other studies related to the use of artificial intelligence technology for milk prediction; Finally, the conclusion is given.

2. Materials and methods or calculation method

This study adopts the methods of comprehensive analysis, investigation and experiment. The comprehensive analysis method allows the analysis and synthesis of literature related to the prediction and decision-making of the dairy industry. Through the questionnaire survey, the evaluation and judgment of experts on the application of analytic hierarchy process are obtained. A case study on the application of analytic hierarchy process is carried out by experimental method, in order to select the most suitable artificial intelligence technology for milk yield prediction. The procedure proposed in its study[36]was adopted as the method of implementing AHP. F**igure 1** depicts this process.

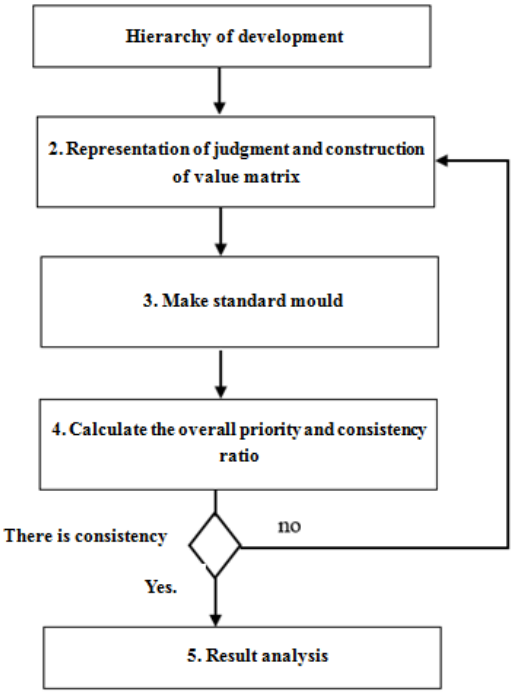

Figure 1. Methodology for implementing AHP[36].

15 national experts with high professional level and experience in artificial intelligence (100% experts) and animal husbandry (80% experts) compared the alternatives according to the selection criteria. Experts are selected through curriculum analysis, taking into account their university degrees, scientific categories, academic publications, mastery of relevant topics and research experience. All experts are doctors of science, and 66.67% of the experts have more than 10 years of research experience. The experts consulted were from the following institutions: Gene company "Camilo Cienfuegos" (2 experts), Universidad de Pinar del Río (2 experts), Provincial Meteorological Center of Pinar del Río (1 expert), Havana Agricultural University (4 experts), Institute of Agricultural Engineering (4 experts) and Universidad de Camagüey (2 experts).

The correct selection of experts is helpful to reduce the error and uncertainty in the process of scheme comparison. In order to prioritize options based on the selected comparison criteria and to improve the certainty of the process, the scale described in **Table 1**[37]was used. The scale allows a degree of certainty and homogeneity to determine the importance or preference of alternatives in the comparison matrix[36].

Table 1**.** The scale used to measure expert judgment[37].

Numerical scale	Speech scale	Description of
1	Equally important	These two elements contribute equally to ownership or standards
3	This factor is more important than another factor.	Judgment and experience tend to one factor rather than another
5	Element is more important than another element	Previous judgments and experiences strongly favor one factor over another
7	The importance of an element is strong relative to another element	Element strong domination
9	The importance of an element is extreme relative to another element	One element dominates another by the highest possible order of magnitude
2,4,6,8	Intermediate value between two adjacent judgments	

3. Results and discussion

3.1. Development hierarchy

Figure 2 depicts the hierarchical analysis performed in this study. At the top of the hierarchy are analysis objectives, comparison criteria at the middle level and alternatives at the lower level.

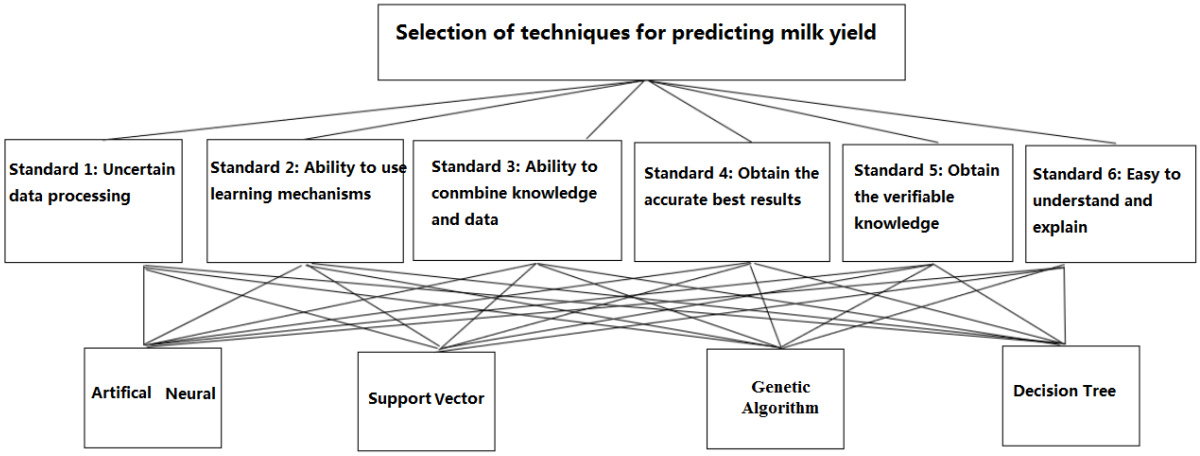

Figure 2. Hierarchy analysis.

Source: self compiled.

These criteria were selected based on a review of the literature on livestock forecasting: the processing of uncertain data[4,16], the use of learning mechanisms[12], the combination of knowledge and data[12], the best accurate results[8,9,29], Gain verifiable knowledge[12], easy to understand and interpret[27].

The alternatives evaluated in this study consist of AI technologies identified in the references. According to[27], these algorithms are not listed as alternative algorithms based on regression, because in recent years, they have only been used as comparative methods in milk production management research, indicating that the yield is lower than other AI technologies. The selected alternatives bring together different algorithms classified according to the following criteria[38]and[27]:

1. ANN-based techniques*: adaptive neuro fuzzy inference system,* symmetrical artificial neural network, *nonlinear autoregressive model with external input,* multilayer perceptron, *back propagation neural network, convolution neural network and long-term and short-term memory network* (LSTM).
2. SVM-based techniques: *support vector regression.*
3. GA-based techniques: simple classical genetic algorithm.
4. DT-based technology: random forest, statistical decision tree, regression tree and classification.

3.2. Representation of judgment and construction of value matrix

Through the questionnaire survey of the expert group, the expert group can make value judgments on the selection criteria and options. Seven n × n pair comparison matrices were prepared, and the experts evaluated the selection criteria and alternatives of assumptions based on these matrices. **Table 2** shows the value matrix based on expert evaluation.

In each matrix, the elements of row i = 1, 2,..., n are rated according to the proportional value described in **Table 1** relative to the elements of column j = 1, 2,..., n. This process is performed by aji = 1/K if the AIJ element of the comparison matrix A is k, and then aii = 1 for all diagonal elements because they are self-evaluated[37]. The geometric mean is used to synthesize the consensus judgment of experts because it provides sufficient accuracy in this process[36].

Table 2. Consensus evaluation matrix

a. Standard comparison matrix

	Standard 1	Standard 2	Standard 3	Standard 4	Standard 5	Standard 6
Standard 1	1	2	3	2	4	2
Standard 2	1/2	1	5	1/3	1/3	1/3
Standard 3	1/4	1/5	1	1/7	1/3	1/5
Standard 4	1/2	3	7	1	5	2
Standard 5	1/3	3	3	1/5	1	1/2
Standard 6	1/2	3	5	1/2	2	1

b. Comparison matrix 1

	ANN	SVM	GA	DT
ANN	1	3	8	3
SVM	1/3	1	5	1/3
GA	1/8	1/5	1	1/6
DT	1/3	3	6	1

c. Standard 2 Comparison Matrix

	ANN	SVM	GA	DT
ANN	1	6	9	8
SVM	1/6	1	4	5
GA	1/9	1/4	1	2
DT	1/8	1/5	1/2	1

d. Standard comparison matrix 3

	ANN	SVM	GA	DT
ANN	1	2	5	2
SVM	1/2	1	8	1/2
GA	1/5	1/8	1	1/7
DT	1/2	2	7	1

e. Standard comparison matrix 4

	ANN	SVM	GA	DT
ANN	1	1/2	4	5
SVM	2	1	3	4
GA	1/4	1/3	1	1/2
DT	1/5	1/4	2	1

f. Standard comparison matrix 5

	ANN	SVM	GA	DT
ANN	1	3	8	2
SVM	1/3	1	6	1/3
GA	1/8	1/6	1	1/4
DT	1/2	3	4	1

.g. Standard comparison matrix 6

	ANN	SVM	GA	DT
ANN	1	1/2	1/6	1/7
SVM	2	1	1/6	1/7
GA	6	6	1	1/3
DT	7	7	3	1

3.3. Manufacture of standardized mould

To standardize the comparison matrix, divide the elements of each comparison matrix by the sum of the values in the corresponding column. The priority vector is calculated by averaging each row of the normalization matrix. **Table 3** shows the normalization matrix and its respective priority vectors.

The priority vector represents the preference of the alternative relative to the standard under consideration. The results in **Table 3a** show that the ability to process uncertain data and obtain the best accurate results are the most relevant criteria for selecting the AI technology most suitable for milk

production prediction. According to these standards, it is found that artificial neural network is more effective in dealing with uncertain data (**Table 3b**), support vector machine obtains accurate results while saving time and computing resources, followed by artificial neural network (**Table 3e**). In addition, there is evidence that DTS is an artificial intelligence technology that provides a better understanding and interpretation of its functions in the prediction process (**Table 3g**).

Table 3. Normalization matrix and its priority vector

a. Standardization matrix and priority vector comparison standard

	Standard 1	Standard 2	Standard 3	Standard 4	Standard 5	Standard 6	Priority vector
Standard 1	0.32	0.16	0.13	0.48	0.32	0.33	0.29
Standard 2	0.16	0.08	0.208	0.08	0.03	0.06	0.10
Standard 3	0.08	0.02	0.04	0.03	0.03	0.03	0.04
Standard 4	0.16	0.25	0.29	0.24	0.39	0.33	0.28
Standard 5	0.11	0.25	0.13	0.05	0.08	0.08	0.11
Standard 6	0.16	0.25	0.21	0.12	0.16	0.17	0.18
Total	1	1	1	1	1	1	1

b. Normalized matrix and priority vector of standard 1

	ANN	SVM	GA	DT	**Priority vector**
ANN	0.558	0.417	0.400	0.667	0.51
SVMe	0.186	0.139	0.250	0.074	0.16
GA	0.070	0.028	0.050	0.037	0.05
DT	0.186	0.417	0.300	0.222	0.28
Total	1	1	1	1	1

c. Normalized matrix and priority vector of standard 2

	ANN	SVM	GA	DT	**Priority vector**
ANN	0.713	0.805	0.621	0.500	0.66
SVM	0.119	0.134	0.276	0.313	0.21
GA	0.079	0.034	0.069	0.125	0.08
DT	0.089	0.027	0.034	0.063	0.05
Total	1	1	1	1	1

d. Normalized matrix and priority vector of standard 3

	ANN	SVM	GA	DT	**Priority vector**
ANN	0.455	0.390	0.238	0.549	0.41
SVM	0.227	0.195	0.381	0.137	0.24
GA	0.091	0.024	0.048	0.039	0.05
DT	0.227	0.390	0.333	0.275	0.31
Total	1	1	1	1	1

e. Normalized matrix and priority vector of standard 4

	ANN	SVM	GA	DT	**Priority vector**
ANN	0.290	0.240	0.400	0.476	0.35
SVM	0.580	0.480	0.300	0.381	0.44
GA	0.072	0.160	0.100	0.048	0.10
DT	0.058	0.120	0.200	0.095	0.12
Total	1	1	1	1	1

f. Standardized matrix and vector priority criterion 5

	ANN	SVM	GA	DT	**Priority vector**
ANN	0.511	0.419	0.421	0.558	0.48
SVM	0.170	0.140	0.316	0.093	0.18
GA	0.064	0.023	0.053	0.070	0.05
DT	0.255	0.419	0.211	0.279	0.29
Total	1	1	1	1	1

g. Normalized matrix and priority vector 6

	ANN	SVM	GA	DT	**Priority vector**
ANN	0.063	0.034	0.038	0.088	0.06
SVM	0.125	0.069	0.038	0.088	0.08
GA	0.375	0.414	0.231	0.206	0.31
DT	0.438	0.483	0.692	0.618	0.56
Total	1	1	1	1	1

Table 4. Priority matrix

	Standard 1	**Standard 2**	**Standard 3**	**Standard 4**	**Standard 5**	**Standard 6**	**Priority vector**
ANN	0.51	0.66	0.41	0.35	0.48	0.06	0.39
SVM	0.16	0.21	0.24	0.44	0.18	0.08	0.23
GA	0.05	0.08	0.05	0.10	0.05	0.31	0.11
DT	0.28	0.05	0.31	0.12	0.29	0.56	0.26

3.4. Calculation of global priority and consistency relation vector

According to the standards listed in **Tables 3b**, **3c**, **3d**, **3e**, **3f** and **3g** respectively, a priority matrix is established, which contains the priority vector of the alternative and is used to calculate the global priority vector of the alternative. **Table 4** shows the constructed priority matrix and the global priority vector of alternatives.

The overall priority vector of alternatives represents their preference level and constitutes a solution judged by experts[36]. In order to determine its value, the standard priority vector obtained in **Table**

3a is multiplied by the priority matrix listed in **Table 4**. The global priority vector shows that the technology based on artificial neural network and decision tree constitutes an artificial intelligence technology with better fitting for milk yield prediction with preference index of 0.39 and 0.26 respectively.

In order to verify the results of this study, we calculated the consistency ratio of each comparison matrix. RC represents the reasonableness of the judgment used in the comparison matrix and is calculated by dividing the consistency index (CI) by the random consistency index (RI)[36]. Formulas 1 and 2 illustrate how to determine the size of these indexes respectively.

$$IC = \frac{\lambda \max - n}{n-1} \quad (1)$$

$$IA = \frac{1.98(n-2)}{n} \quad (2)$$

Where

n: compare the number of elements in the matrix to be evaluated.

λmax: the average value of vector elements, which is obtained by multiplying the comparison matrix by its respective priority vector and the product of the latter.

According to Saaty[39] , if RC < 0.10, the degree of inconsistency in judgment is acceptable, otherwise experts must reassess their evaluation. **Table 5** shows the RC values of the comparison matrix.

Table 5. Consistency relationship of comparison matrix

Matrix comparison	Reinforced
Comparison standard	0.091
Uncertain data processing	0.067
Ability to use learning mechanism	0.079
Ability to combine knowledge and data	0.078
Get accurate results in the best way	0.079
Gain verifiable knowledge	0.075
Easy to understand and explain	0.060

The results in **Table 5** show that the degree of inconsistency in peer comparison is acceptable, which indicates that there is no contradiction in the judgment of experts.

4. Result analysis

The RC index described in **Table 5** can infer that the execution of the expert judgment process is correct. The results show that the artificial intelligence technology based on artificial neural network is more suitable for predicting milk yield with 39% preference level than the technology based on decision tree and support vector machine (26% and 23% respectively).

In their analysis[29], they found that NARX artificial neural network can improve the accuracy of milk yield prediction better than the model based on multiple linear regression and static artificial neural network technology. The author[38]determined that the most used artificial intelligence technologies in animal husbandry production related research during 2004–2018 were support vector machine and artificial neural network. In their study[8], they compared the performance of three prediction models to predict the milk yield of Holstein Frisian varieties using neural network, support vector machine and random forest techniques, respectively. The experimental results show that the model based on support vector machine has high accuracy and computational complexity. According to[19], using GA in LSTM network can obtain more accurate milk yield prediction than the model based on ANN-LSTM only[27]. A retrospective study of dairy farm management from 2010 to 2020 shows that artificial neural network and DT are the best performing artificial intelligence technologies in this field. Consistent with the research results of Slob, Catal and Kassahun[26], this study shows that the artificial intelligence technology based on artificial neural network and DT provides the best performance for predicting milk yield respectively.

5. Conclusion

With the application of analytic hierarchy process, the technology based on artificial neural network is considered to be the most suitable technology for milk yield prediction, which is better than the technology based on decision tree and support vector machine. In addition, there is evidence that the ability to process uncertain data and obtain accurate results in the best way are the most relevant selection

criteria for evaluating AI technologies applied to milk yield prediction.

This study is helpful to the decision-making of milk production organization and the development of prediction model of dairy industry. Future work can be used to evaluate the performance of different types of artificial neural networks in predicting milk yield and determine the correlation between their characteristics and their efficiency as a prediction tool.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Wen Z, Liao H, Zavadskas EK. Macont: Mixed sggregation by vomprehensive normalization Technique for multi-criteria analysis. Informatica 2020; 31(4): 857–880.
2. Mahmoudi A, Mi X; Liao H, et al. Grey best-worst method for multiple experts multiple criteria decision making under uncertainty. Informatica 2020; 31(2): 331–357.
3. Gunnar B. Different Methods to forecast milk delivery to dairy: A comparison for forecasting. International Journal of Agricultural Management 2015; 4(3): 132–140.
4. Yan WJ, Chen X, Akcan O, et al. Big data analytics for empowering milk yield prediction in dairy supply chains. In: International Conference on Big Data (Big Data). Santa Clara, CA, USA: IEEE; 2015. p. 2132–2137.
5. Zhang F, Upton J, Shalloo L, et al. Effect of parity weighting on milk production forecast models. Computers and Electronics in Agriculture 2019; 157: 589–603.
6. Jensen DB, Van Der Voort M, Hogeveen H. Dynamic forecasting of individual cow milk yield in automatic milking systems. Journal of Dairy Science 2018; 101(11): 1–12.
7. Da Rosa R, Goldschmidt G, Kunst R, et al. Towards combining data prediction and internet of things to manage milk production on dairy cows. Computers and Electronics in Agriculture 2020; 169.
8. Nguyen QT, Fouchereau R, Frénod E, et al. Comparison of forecast models of production of dairy cows combining animal and diet parameters. Computers and Electronics in Agriculture 2020; 170.
9. Perdigón R, González N. A literature review on models to predict milk yields. Revista Ingeniería Agrícola 2020; 10(4): 69–77.
10. Torres-Inga CS, López-Crespo G, Guevara-Viera R, et al. Technical efficiency in dairy farms in the sierra andina through modeling with neural networks. Revista Producción Animal 2019; 31(1): 10–15.
11. 11. Kaygisi F, Sezgin FH. Forecasting Goat Milk production in Turkey using artificial neural networks and box-jenkins models. Animal Review 2017; 4(3): 45–52.
12. González N, Leiva MY, Faggioni KM, et al. Comparative study of Artificial Intelligence techniques for the diagnosis of diseases in livestock. Revista de Sistemas, Cibernética e Informática 2018; 15(2): 17–20.
13. 13. Sugiono S, Soenoko R, Riawati L. Investigating the impact of physiological aspect on cow milk production using artificial intelligence. International Review of Mechanical Engineering 2017; 11(1): 30–36.
14. Gorgulu O. Prediction of 305 days milk yield from early records in dairy cattle using on fuzzy inference system. The Journal of Animal & Plant Sciences 2018; 28(4): 996–1001.
15. Machado G, Figueiredoa DM, Resende PC, et al. Predicting first test day milk yield of dairy heifers. Computers and Electronics in Agriculture 2019; 166: 1–8.
16. Liseunea A, Salamone M, Van Den Poel D, et al. Leveraging latent representations for milk yield prediction and interpolation using deep learning. Computers and Electronics in Agriculture 2020; 175:105600.
17. Zhang F, Murphy MD, Shalloo L, et al. An automatic model configuration and optimization system for milk production forecasting. Computers and Electronics in Agriculture 2016; 128: 100–111.
18. Zhang F, Upton J, Shalloo L, et al. Effect of parity weighting on milk production forecast models. Computers and Electronics in Agriculture 2019; 157: 589–603.
19. Zhang W, Yang K, Yu N, et al. Daily milk yield

prediction of dairy cows based on the GA-LSTM Algorithm. In: International Conference on Signal Processing (ICSP). Beijing, China: IEEE; 2020. p. 664–668.

20. Saha A, Bhattacharyya S. Artificial insemination for milk production in India: A Statistical Insight. Indian Journal of Animal Sciences 2020; 90(8): 1186–1190.
21. Notte G, Pedemonte M, Cancela H, et al. Resource allocation in pastoral dairy production systems: Evaluating exact and genetic algorithms approaches. Agricultural Systems 2016; 148: 114–123.
22. Eyduran E, Yilmaz I, Tariq M, et al. Estimation of 305-d milk yield using regression tree method in Brown Swiss Cattle. JAPS Journal of Animal and Plant Sciences, 2013, 23(3): p. 731– 735.
23. 23. Piwczyński D, Sitkowska B, Kolenda M, et al. Forecasting the milk yield of cows on farms equipped with automatic milking system with the use of decision trees. Animal Science Journal 2020; 91(1): e13414.
24. Kliś P, Piwczyński D, Sawa A, et al. Prediction of lactational milk yield of cows based on data recorded by AMS during the periparturient period. Animals 2021; 11(2):383.
25. Dongre VB, Gandhi RS, Singh A, et al. Comparative efficiency of artificial neural networks and multiple linear regression analysis for prediction of first lactation 305-day milk yield in Sahiwal Cattle. Livestock Science 2012; 147(1–3): 192–197.
26. Oscullo J, Haro L. Forecast of the daily demand of the national interconnected system using neural networks. Revista Politécnica 2016; 38 (1): 77–82.
27. Slob N, Catal C, Kassahun A. Application of machine learning to improve dairy farm management: A systematic literature review. Preventive Veterinary Medicine 2021; 187: 105237.
28. Perdigón R, Viltres H, Orellana A. Models for predicting perishable products demands in food trading companies. Revista Cubana de Ciencias Informáticas 2020; 14(1): 110–135.
29. Murphy MD, O'mahony MJ, Shalloo L, et al. Comparison of modeling techniques for milk-production forecasting. Journal of Dairy Science 2014; 97(6): 3352–3363.
30. González N, Estrada V, Febles A. Study and selection of artificial intelligence techniques for the diagnosis of diseases. Revista de Ciencias Médicas de Pinar del Río 2018, 22(3): 534–544.
31. Saaty TL, Ergu D. When is a decision-making method trustworthy? Criteria for rvaluating multi-vriteria decision-making methods. International Journal of Information Technology & Decision Making 2015; 14(6): 1171–1187.
32. Ossadnik W, Schinke S, Kaspar RH. Group aggregation techniques for analytic hierarchy process and analytic network process: A comparative analysis. Group Decision and Negotiation 2016; 25: 421–457.
33. Rozman Č, Grgić Z, Maksimović A, et al. Multiple-criteria approach of evaluation of milk farm models in bosnia and herzegovina. Mljekarstvo 2016; 66(3): 206–214.
34. Montalván-Estrada A, Aguilera-Corrales Y, Veitia-Rodríguez E, et al. Multicriteria analysis for theIntegrated management of industrial wastewater. Ingeniería Industrial 2017; 38(2): 56–67.
35. Ocampo CD, Tamayo J, Castaño HM. Risk management in the implementation of photovoltaic systems in gold extraction projects in colombia from the Hierarchical Analysis Process (AHP). Información Tecnológica 2019; 30(3): 127–136.
36. Mendoza A, Solano C, Palencia D, et al. Application of the Analytical Hierarchy Process (AHP) for decision making with expert judgments. Ingeniare. Revista chilena de ingeniería 2019; 27(3): 348–360.
37. Saaty TL. How to Make A Decision: The analytic hierarchy process. European Journal of Operational Research 1990; 48(1): 926.
38. Liakos KG, Busato P, Moshou D, et al. Machine learning in agriculture: A Review. Sensors 2018; 18: 2674.

Journal of Autonomous Intelligence (2021) Volume 4 Issue 2.
doi: 10.32629/jai.v4i2.503

Original Research Article

Application of Artificial Intelligence Technology in Automatic Detection of large Intestine Polyps

Martin Alonso Gomez-Zuleta[1], Diego Fernando Cano-Rosales[2]*, Diego Fernando Bravo-Higuera[3], Josue Andre Ruano-Balseca[4], Eduardo Romero-Castro[5]

[1] *Gastroenterology and Echoendoscopy Unit. National University Hospital. National University of Colombia, Colombia*

[2] *Gastroenterology, National University of Colombia. National University Hospital. Bogota D.C., Colombia*

[3] *Militar University of New Granada. National University of Colombia. Bogota D.C., Colombia*

[4] *Systems and computing. National University of Colombia. Bogota D.C., Colombia*

[5] *Biomedical Sciences, National University of Colombia. Bogota D.C., Colombia*

ABSTRACT

Objective: to establish an automatic colonoscopy method based on artificial intelligence. **Methods:** a public database established by a university hospital was used, including colorectal fat and data collection. Initially, all frames in the video are normalized to reduce the high variability between databases. Then, the convolution neural network is used for full depth learning to complete the detection task of polyps. The network starts with the weights learned from millions of natural images in the ImageNet database. According to the fine-tuning technology, the colonoscopy image is used to update the network weight. Finally, the detection of polyps is performed by assigning the probability of containing Po´ lipo to each table and determining the threshold defined when polyps appears in the table. **Results:** 1875 cases were collected from 5 public databases and databases established by university hospitals, with a total of 123046 forms. The method was trained and evaluated. Comparing the results with the scores of different colonoscopy experts, the accuracy was 0.77, the sensitivity was 0.89, the specificity was 0.71, and the ROC curve (re ceiver operation characteristics) was 0.87. **Conclusion:** compared with experienced gastrointestinal markers, this method overcomes the high variability of different types of lesions and different colonic light conditions (handle, folding or contraction), has very high sensitivity, and can reduce human errors, which is one of the main factors leading to the non detection or leakage of Po lipids in colonoscopy.

Keywords: *Machine Colonoscopy; Colorectal cancer; Polyp; Screening; Artificial intelligence*

ARTICLE INFO

Received: Sep 28, 2021
Accepted: Nov 16, 2021
Available online: Nov 20, 2021

***CORRESPONDING AUTHOR**

Diego Fernando Cano-Rosales
dfcanor@unal.edu.co;

CITATION

Gomez-Zuleta MA, Cano-Rosales DF, Bravo-Higuera DF, *et al*. Application of artificial intelligence technology in automatic detection of large Intestine Polyps. Journal of Autonomous Intelligence 2021; 4(2): 72-81. doi: 10.32629/jai.v4i2.503

1. Introductory

Colorectal cancer is the third largest cancer in the world and the second leading cause of cancer death. In Colombia, it is the fourth most common tumor in men and women, with a higher incidence that increase every year[1,2]. Many studies have concluded that screening for RCC is cost-effective in people at moderate risk (people without family history and susceptibility History). It is well known that age (≥50 years), eating habits and smoking are risk factors for increasing the incidence of this disease. In the general population, the risk is 5%–6%, and the incidence rises sharply from the age of 50. Therefore, people aged 50 or over are considered to be at medium risk, and the screening scheme should be started[3,4].

As for the survival rate of RCC patients, it is directly related to the severity of the disease at the time of diagnosis. Individuals diagnosed with advanced stage had a survival rate of 7% within 5 years, while

92%[5] of patients with early RCC were reported; Therefore, early detection of tumors, or more precisely, detection of adenomatous (precancerous) polyps, so as to prevent disease, is of great significance. It is well known that using existing screening techniques (occult blood, colonoscopy), RCC can be highly prevented in more than 90% of cases.

Many studies have shown that colonoscopy is the first choice for the prevention and early detection of RCC because, as mentioned above, it can detect the main sources of RCC, such as adenomatous polyps[6-9].

In addition to early detection of cancer, if timely treatment is completely curable, polyp detection is a quality indicator of colonoscopy, and 20% of women and 30% of men are considered to be an indicator of adenomatous polyps (with high cancer risk) during the examination; This means that an average of 25% of colonoscopy should find adenomatous polyps. Unfortunately, different studies report that about 26% of polyps are not found during colonoscopy, which is a very high error rate, which is mainly explained by two factors: the number of blind spots in colonoscopy (lipid behind folds, colon stalk, preparation, etc.) And human errors ience affecting the smoothness of colon scanning. At present, some strategies have solved this challenge as a classification task and used machine learning technology.

On the one hand, some authors try to screen liposome candidates from low-level characteristics. Bernal et al.[13] proposed an appearance model of polyps, which describes the polyps valleys as a continuous and concave boundaries. This feature is used to train the classifier, which obtains a polyps detection sensitivity of 0.89 in a test set (test). Shin et al.[14] proposed a block by block classification strategy, using the combination of shape and color features to obtain a sensitivity of 0.86. On the other hand, some studies use deep convolutional neural network (CNN), which is a set of rhythms gathered at the end of *deep learning*. Urban et al.[15] proposed a convolution network for real-time detection of liposomes of different sizes, with a sensitivity of 0.95. However, Taha and cola borators[16] analyzed some limitations of these works, one of which is that these works related to surgery[10-12]. A lot of work has been done to try to solve these two factors to reduce the rate of fat loss, which is why people have designed some accessories to enable the fat hidden behind the folds to be identified, such as hats, cuffs, and even a micro endoscope called the "*third eye"*, which tries to flatten the folds or see the back of the folds. In addition, it has recently been considered that the factors related to human errors can be mitigated at least by introducing a second reader (computers). In this case, technology and artificial intelligence begin to display results, which can significantly improve the detection rate of polyps and allow the number of undetected polyps in gastrointestinal units to be reduced.

Developing computational strategies for pattern extraction and automatic detection of colorectal fat in colonoscopy video is a very complex problem. The video of colonoscopy is recorded in a large number of noise sources, which are easy to cover up the lesions; For example, the gloss on the intestinal wall caused by reflective light source or specular reflection, the movement of organs and intestinal secretions blocking the field of view of colonoscopy, and expert exper

require a lot of data for training. In addition, these databases are obtained under specific climatic conditions; In particular, imaging equipment, scanning protocols executed by experts and the extraction of sequences vulnerable to visual damage. Although some progress has been made, there is still a challenge to develop generalizable models to accurately detect lesions, regardless of the type of injury, the form of expert scanning or the colonoscopy unit used.

The main purpose of this study is to establish a strategy for automatically detecting colorectal polyps. It is suggested to construct a second reader to support the colon scanning process and reduce the number of undetected lesions in the process of settler replication. This paper presents an automatic classification strategy of polyps in colonoscopy video sequence. This study is based on a rhythmic in-depth learning and evaluates the different architectures of convolutional networks. The tissue structure of this paper is as follows: firstly, the automatic detection

method of polyps is introduced; then, it describes the policy considerations surrounding this work; then, the experimental configuration is displayed together with all test results and compared with expert notes; then it introduces the discussion of this paper; Finally, the conclusion is drawn and the future work is put forward.

2. Methodology

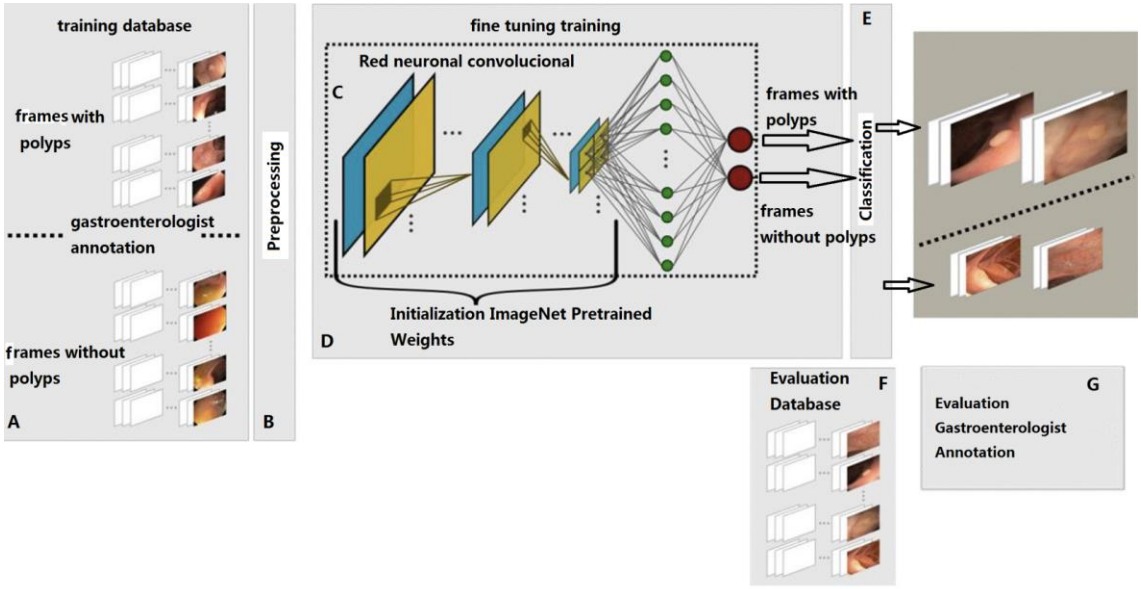

Figure 1. Method for automatic detection of polyps. First, it integrates an annotated colonoscopy video database (**A**). Each of these tables is preprocessed (**B**) to feed back some CNN based models (**C**). This model is fine tuned and trained with the preset weights of millions of natural genes (**D**). In the trained network, the test database (**F**) was used **to** evaluate its performance in detecting polyps**(H),** and the results were compared with expert note (**G**).

This paper introduces a deep learning method that simulates the high variability in the process of colonoscopy. This method aims to automatically detect the lipids in the process of colonoscopy. This task is divided into two stages: training and classification. Firstly, frame-by-frame preprocessing is performed, common to both stages the two eta-pas are preprocessed table by table and generally. Then, convolutional neural networks were trained using a large number of colonoscopy images annotated by colonoscopy experts (with about 20 years of experience and 50 thousand000 colonoscopy examinations), which were divided into two categories: negative class or without polyps *Po´* lipo, positive class or with polyps*Po´* lipo. The model obtained from the learning process is used to classify new genes (or not used in the naming process) into one of these two categories. The process of this work is shown in **Figure 1**, as follows.

3. Acquisition protocol and preprocessing

In order to reduce the influence of various noise sources on different colonoscopy acquisition processes and colorectal physiological conditions, it is necessary to preprocess the video frame by frame. Firstly, each frame is normalized with an average value of 0 and a standard deviation(SD) of 1, so that the features extracted between frames are comparable. Then, capture presents different spatial resolutions according to different shooting devices, so each frame is scaled to 300 x 300 pixels so that each frame has the same capture grid.

4. CNN architecture

The main unit of these structures is neurons, which provide output as a function of input. The arrangement of neurons forms a layer or block, and the network is composed of multiple basic blocks. These basic blocks are arranged in the following

order: multiple pairs of layers with volume (**Figure. 1C**, blue table) and clustering layers (**Figure. 1C**, yellow table), which provide the feature vector of the image, Then there is a set of fully connected layers **(Figure 1C**, green circles), which is used to calculate the probability that a set of features belong to a certain class, and ends with the activation layer (**Figure 1C**,red circles**).** In the activation layer, the obtained probability is normalized and the required binary classification is obtained. The functions of these blocks are:

- Convolution layers: recognize the local features in the whole image, such as the pattern of shape, edge and texture, which is very important to describe the image. This layer connects a subset of adjacent pixels of an image or neuron to all nodes of the first convolution layer. One of the convolution layers or cores is distinguished by the specific weight of each node; by operating on a specific area of the image, it provides a feature map of the area.
- Pooling layers: reduce the computational complexity, so as to reduce the amount of features in the convolution layer, and obtain the hierarchical set of image feature mapping.
- Fully-connected layers: This layer connects each of the neurons in the previous layer to each of the neurons in the next layer. The previous layer is a flat or vector representation of the obtained feature maps. The number of neurons in the next layer is determined by the number of classes that are required to be classified. Finally, the fully connected layer provides a vote
- Activation function: normalizes the probabilities obtained from the completely connected layers according to a specific function, in which a probability of 0 to 1 is obtained.

A particular architecture is made up of an array of modules that contain different configurations and orders of the fundamental blocks explained above, and the result obtained by each neuron is known as a gradient. Three architectures highly evaluated and validated in the state of the art were used in this work: InceptionV3, Vgg16 and ResNet50. Next, we describe each one of them.

- Inception V3: composed of 48 floors, 24 million feet. To a large extent, these layers are divided into 11 levels, and the features of multiple levels are extracted on these levels. Each module consists of a given convolution and grouping layer structure, which is rectified by *linear rectifier unit* (ReLu) function. It ends with an activation function called *exponential normalization* (softmax)[17].
- Vgg16: 16 floors in total, 138,000 meters. 13 layers are convoluted, one layer is clustered (in some cases), the two layers are completely connected, and finally the normalized exponential activation function. This structure is famous for using a small filter of size 3 x 3 on the convolution layer. It has lower computational costs than most architectures[18].
- RESNET 50: composed of 50 floors, 26 million feet. This structure is constructed under the concept of residual network. Obviously, in such a very deep area, the gradient of propagation will disappear in the last layer. To verify this, some layers are trained using the gradient residuals obtained at this point and the gradients at the previous two positions. This structure ends with a standardized exponential activation function[19].

5. Fine tuning training

The high performance of classification mainly depends on the number of IMA genes and how to start training the weight of CNN. Colonoscopy has about 12,000 frames per video, so the availability of databases with IMA gene annotation is limited. Then, the limited data is used for training, and the network weight is started randomly as usual, resulting in the failure of the training process. In order to avoid this uncertainty, we use the weights of the same type of networks (*transfer learning*), which have previously been trained for another natural gene classification problem, and the database contains a large number of unprocessed genes. The reason for this is that even if natural genes and genes in colonoscopy are different, their statistical structures are similar,

and they are composed of original genes representing objects. In this case, networks trained to recognize objects in natural genes are used as the initial conditions for training these networks to recognize polyps.

The use of these weights is achieved through a process called fine tuning. For this purpose, the whole pre focus network is removed and the last fully connected layer is removed. This layer is replaced by a new layer, which has the same number of neurons as the number of classes in the classification task (polyp-on polyp) and starts from the weight of the pretained network. Then, the last layer is trained first, and then the weights of the other layers of the network are updated in the iterative process; this approach is known as backpropagation. Each iteration of this training is carried out using a certain number of samples or batches of the training images. At the end of this process, all sample sets of the network are trained, which is called the training epoch. Determine the number of epochs according to the complexity of the samples to be classified. Finally, when the probability of a training image is high and consistent with the labeled label, the training reaches a climax.

6. Polyp detection

Using the trained network model, it is applicable to a set of evaluation videos, in which a label is classified and assigned: (1) Tables with and (0) without the presence of polyps. However, there are also some surfaces with structures similar to fat, such as burrs produced by intestinal fluids. In these tables, the model shows a classification error and regards this table as a lesson. By temporarily analyzing these errors, it is worth noting that they are displayed as outliers (3 to 10 frames) in a small window of time (60 frames or 2 seconds). Therefore, the classification conducted by the network is temporarily filtered, and it is determined that if at least 50% of the 60 consecutive tables are classified without polyps, the remaining four tables are filtered and assigned a new label, such as the table without polyps. Finally, a polyp is detected when the proposed method classifies an image as a box with a polyp present or a positive class.

7. Data base

The database construction in this study was designed to capture the maximum variability of colonoscopy pretreatment. In order to train and evaluate the proposed method, we collected sequences from different gastrointestinal centers, which contain different types of polymorphic and non polymorphic lesions (morphology and colon location), scanning and capture devices of different experts. The following are the details of these databases.

7.1. ASU Mayo clinical colonoscopy video database

The equipment was built in the gastroenterology department of Mayo Clinic in Arizona. It consists of 20 colonoscopy sequences, which are divided into 10 sequences with polyps and 10 sequences without polyps. These notes were performed by gastroenterology students and verified by experts. This collection is widely used in the latest technology and serves as a database for the “2015 ISBI Grand Challenge on Automatic Polyp Detection in Colonoscopy Videos”[20].

7.2. CVC-ColonDB

It consists of short sequences of 15 different lesions, with a total of 300 frames. This set of lesions is highly variable and difficult to detect because they are very similar to healthy areas. Each painting has an expert gastrointestinal sign. The church was built at the Clinical Hospital of Barcelona, Spain[13].

7.3. CVC-ClinicDB

It consists of 29 short sequences with different injuries that bring together 612 frames annotated by an expert. This database was used by the MICCAI 2015 Sub-Challenge on Automatic event training set Polyp Detection Challenge in Colonoscopy Videos. This collection was built at the Clinical Hospital of Barcelona, Spain[21].

7.4. ETISLarib Polyp DB

It has 196 images, and each gene has an expert annotation. The database is used for the test set of MICCAI 2015 Sub-Challenge on Automatic Polyp Detection Challenge in Colonoscopy Videos[22].

7.5. Kvasir dataset

It is a database that was collected using endoscopic equipment at Vestre Viken Health Trust (VV), in Norway. The images are annotated by one or more medical experts from VV and the Norwegian Cancer Registry (CRN). The data set consists of the images with different resolution from 720 x 576 up to 1920 x 1072 pixels[20].

7.6. HU-DB

This collection was built at the University Hospital of Bogota. There are 253 colonoscopy videos, a total of 233 lesions. Each frame was recorded by a colonoscopy expert who had about 20 years of experience and performed 50,000 colonoscopies.

Each video is shot at 30 frames per second with spatial resolutions of 895 x 718, 574 x 480 and 583 x 457. A total of 1875 cases, 48,573 cases of polyps and 74,548 cases without polyps were integrated into the case database. Each frame of these videos was rated positive by experts. If there is fat, it is positive. If there is no fat, it is negative. **Table 1** summarizes the number of videos and tables for each database used in this study.

Table 1. Describe the number of colonoscopy videos or cases and video frames in each database used in this study*

Data base	Video only		Surface	
	polyp	Polyp free	polyp	Polyp free
ASU-Mayo	10	10	4683	13,481
CVC-ClinicDB	29	0	612	0
CVC-ColonDB	15	0	379	0
ETIS	28	0	196	0
Kvasir	1000	500	1000	500
HU	233	50	41,70	60,567
Total	1315	560	48,57	74,548

*Merging multiple databases to train and evaluate the proposed method can cover a wide range of damage variability.

8. Policy considerations

This work is carried out in accordance with resolution °008430 of 1993, which sets out the scientific, technical and administrative standards for human research (article 11). The project is classified as the least risky research because it only needs to use digital genes, which are generated from anonymously copied colonoscopies videos; in other words, there is no way to know the name or identity of the research object.

9. Result

The CNN used in this article includes InceptionV3, Rresnet50 and Vgg16. The labels assigned to each network were compared with the notes reassigned by experts in each table. The following experimental configuration and evaluation methods are applied to each of the architecture.

9.1. Experimental configuration

CNN previously was trained with images from the public ImageNet database, which contains about 14 million natural images. The resulting weights are used to start a new colonoscopy numerical training process by fine-tuning method. All this updated the weights and trained the network with the settler replica database. The weight update is carried out in the whole training set, with only 120 epochs. Each epoch will train the model through a batch of 32 frames until it covers all frames. For each network, the decision threshold is set manually to maintain a balance in the classification or unpacking of the two categories. The training program is that 70% of the database is used for training and 30% of the database is used to verify the number of cases; In other words, the data is separated from the beginning, and the training, verification and test data are never mixed. A total of 213 cases (24,668 frames) of polyps and 36 videos (27,534 frames) were trained and verified. The evaluation was performed with 103 videos (23,831 frames) and 25 videos (47,013 frames). The details of this set are shown in **Table 2**.

9.2. Quantitative evaluation

The proposed method automatically detects polyps in colonoscopy video; this task is as binary classification problem. This method sets a label for each frame as either a negative class (a frame that does not contain polyps) or a positive class (a frame that contains polyps). In order to evaluate the performance of this work, it compares the estimated or predicted tags with the tags annotated by experts. This comparison allows the calculation of the confusion matrix, which includes the following:

- True-positives (TP): the number of frames that the model correctly classifies as positive classes.
- True-negatives (TN): the number of frames that the model correctly classifies as negative classes.
- False-positives (FP): the number of frames incorrectly classified as positive classes by the model.
- False-negatives (FN): the number of frames that the model incorrectly classifies as negative classes.

Using the fuzzy matrix, four classification methods are selected and calculated. These methods evaluate the classification results of the tables with (positive class) and without (negative class) polyp, as well as the overall prediction ability of the two categories:

- Sensitivity measures the proportion of the corresponding classification table containing Po lipids.
- The proportion of lipid containing specificity is not calculated correctly.
- The accuracy shows the prediction ability of me all for polyp image classification.

  Accuracy is the ratio of correctly classified tables, depending on the total number of these tables.

The results obtained are given by each of the deep learning architecture explained in section of methodology. The results of each of the architecture are shown in **Table 3**.

Table 2. Describe the number of sequences and frames selected from each database to evaluate the proposed performance or method*

Data base	Video only		Surface	
	Polyp	Polyp free	Polyp	Polyp free
ASU-Mayo	5	2	2124	2553
CVC-ClinicDB	9	0	191	0
CVC-ColonDB	4	0	145	0
ETIS	7	0	45	0
HU	78	23	21,32	44,460
Total	103	25	23 821	47 013

*This accounts for about 30% of the total number of databases.

Table 3. Results of all proposals*

Metric	InceptionV3	Resnet50	Vgg16
Accuracy	0.81	0.77	0.73
Sensitivity	0.82	0.89	0.81
Specificity	0.81	0.71	0.70
Precision	0.67	0.59	0.56
F1 score	0.74	0.71	0.66
ROC (area under the curve)	0.85	0.87	0.81

*The architecture in the evaluation is specified in the column and each method used is specified in the row.

On the one hand, although most of these architectures show excellent performance in classification tasks, Resnet50 architecture has the best performance because I have well detected positive classes or frames of polyp and obtained a sensitivity of 0.89. On the other hand, the structure of InceptionV3 is the best structure for detecting negative class or frames without polyps, with a specificity of 0.81. In order to evaluate the performance of these architectures in detail, ROC (receiver operating characteristic) curves were constructed by each architecture. In this representation, we try to analyze the model and divide images into specificity and sensitivity by changing the decision threshold according to the probability provided by the model. As shown in **Figure 2**, the Resnet50 architecture can better separate classes without considering the decision threshold. This shows that the architecture can better summarize the variability within and between classes.

10. Conclusion

The detection of adenomatous polyps is the main quality index of colonoscopy, because it is a key index for screening and prevention of CCR. In many countries, the quality of gastrointestinal markers is measured by the amount of these polyps, which can be detected in all colonoscopy examinations. The average rotation of experts is about 25%, but for inexperienced gastrointestinal markers, it may be as low as 10%, which leads to the latter escaping more adenoma.

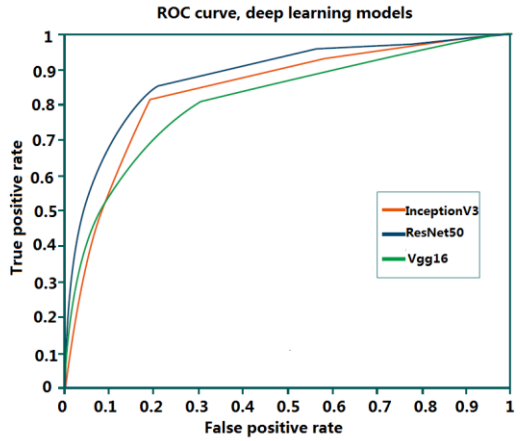

Figure 2. ROC curves for each of the evaluated architectures. The orange line corresponds to the curve of the InceptionV3 architecture; the blue line, to the ResNet50 architecture; and the green line, to Vgg16 architecture. The ResNet50 architecture has better performance, with an area under the curve of 0.87.

This is why some studies[10-12] reported that 26% of polyps were not detected during colonoscopy, which may contribute to more cases of CRC. By 2018, there are 1.8 million new cases worldwide (IARC, 2018)[1]. This is because several factors affect the adequacy of colon scanning, such as the expert's experience and concentration level throughout the working day (related to fatigue), the physiological condition of the colon is a blind spot, and it is difficult to locate the colonoscopy due to the movement of the organ itself and the patient's previous colon preparation, The observability of the colon wall is determined according to the cleanliness of the colon wall[23]. Most of these factors warn that colonoscopy is highly dependent on human factors and requires a second reader unaffected by these factors. In practice, the use of computer tools to detect polyps helps to confirm the findings of experts and, more importantly, remind experts of possible undetected lesions. Therefore, these tools will help to reduce the undetected lipid rate of polyps and thus reduce the incidence of CRC.

In order to support CRC diagnosis using tools of computer vision, this challenge has been solved as follows:

- Detection, referring to the frame-by-frame binary classification of a video into positive class (with polyp) and negative class (without polyp);
- Localization, such as thick delineation (using a box) of the lesion on an image containing polyp;
- Segmentation, as a fine delimitation of the lesion (outlining the edge of the polyp).

The detection of polyps is the primary task facing the whole gastrointestinal tract. The post detection task (localization and segmentation) is a useful process for the expert because he has detected the damage, needs to describe it morphologically, and classifies the graphical user interface as the Paris Classification[6]. This classification enables you to determine the effective management of short-term and long-term diseases. Therefore, these tasks depend entirely on the accuracy of previous detection; therefore, the proposed method completely focuses on the main task required by experts: obtaining colonoscopy images with lesions. In addition, in the prior art, the work of processing these tasks[13-15] describes the limitations when presenting a flow containing at least two tasks. These tasks use different methods for each task, because each task has its own degree of completion. Generally speaking, the polyps frame is detected by measuring the context or global relationship at the image level; the locality and segmentation are analyzed at the pixel level by measuring the local relationship.

In this paper, a robust polyps detection strategy is proposed, which is solved as a classification problem. In the past five years, due to the development of technology, the use of these models has increased to a great extent. These technologies re-

quire a large number of parallel processing and the release of millions of gene databases, such as ImageNet. This makes it possible to design and train highly complex networks, so as to obtain high performance in classification tasks, because it can simulate the high variability of shape, color and texture. However, in the medical field, there is no large amount of public annotation data, so these models are not considered to be applied to disease screening or classification.

The development of transfer learning technology (or transfer learning) provides a solution to solve the shortage of medical data. Use the network weights of millions of natural image training to start a new network and train it with less different data, such as images in colonoscopy. Recent studies using this stream have shown its ability to appropriately summarize the high variability of images with or without polyps lesions in colonoscopy in a specific database. However, different types of lesions and physiological conditions of large intestine are not the only source of variation. The lower the expertise of the specialist, the videos are likely to have a higher number of noisy frames produced by occlusions and abrupt movements of the colonoscopy. In addition, the capture device is also different in light source and visual angle. Therefore, as with existing studies[13-15], training and validation using a database obtained from a specific single gastrointestinal service did not cover all the variability of colonoscopy image classification tasks.

Therefore, in this work, we integrated an attached training video with high variability, which has not appeared in the current technology when collecting security from different databases. This method includes: lesions of different sizes, positions and shapes; colonoscopy and anal surgery by different gastrointestinal experts; and videos taken using different settler reproduction units. Despite this variability, the sensitivity and specificity of detecting polyps in colonoscopy sequence were 0.89 and 0.71.

11. Conclusion

At present, deep learning method is a promising choice, which can be used in classification tasks. With the progress of technology and the continuous design and evaluation of the network, it is possible to integrate a complete set of processes to achieve high performance. Through the evaluation of these networks, the results show that they can be used as the second reader in colonoscopy service.

It is worth noting that these networks fully summarize the high variability of colonoscopy video. The results obtained show that the proposed method can significantly distinguish images regardless of the presence or absence of polyps, regardless of whether the specific clinical protocol of video recording refers to the expert executing the capture program and equipment. This may be an effective method to reduce the detection rate of adenomas for gastrointestinal experts and beginners.

As future work, the proposed method should be tested in a complete colonoscopy procedure and assess whether it is possible to implement it in real time. In addition, a strategy should be developed to not only detect, but also define hazards in the form.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality for 36 cancers in 185 countries. CA Cancer J Clin 2018; 68(6): 394–424.
2. Data I, Method L. Globocan Colombia 2018. 2018; 380: 1–2.
3. Samadder NJ, Curtin K, Tuohy TM, et al. Characteristics of missed or interval colorectal cancer and patient survival: A population-based study. Gastroenterology 2014; 146(4): 950–960.
4. Katenbach T, Sano Y, Friedland S, et al. American gastroenterological association. American Gastroenterological Association (AGA) institute technology assessment on image-enhanced endoscopy. Gastroenterology 2008; 134(1): 327–340.

5. Brown SR, Baraza W, Din S, et al. Chromoscopy versus conventional endoscopy for the detection of polyps in the colon and rectum. Cochrane Database Syst Rev 2016; 4: CD006439.
6. The Paris Endoscopic Classification of Superficial Tumor Lesions: Esophagus, Stomach and Colon: November 30 to December 1, 2002. Gastrointestinal 2003; 58 (6 supplement): s3-43.
7. Dinerson L, Chua TJ, Kaffees AJ. Meta-analysis of narrowband imaging versus conventional colonoscopy for adenoma detection. Gastrointest Endosc 2012; 75(3): 604–611.
8. Nagorni A, Bjelakovic G, Petrovic B. Narrow band imaging versus conventional white light colonoscopy for the detection of colorectal polyps. Cochrane Database Syst Rev 2012; 1: CD008361.
9. Jin X, Chai S, Shi J, et al. Meta-analysis for evaluating the accuracy of endoscopy with narrow band imaging in detecting colorectal adenomas. Journal of Gastroenterology and Hepatology 2012; 27(5): 882–887.
10. Komenda Y, Suzuki N, Sarah M, et al. Factors associated with failed polyp retrieval at screening colonoscopy. Gastrointest Endosc 2013; 77(3): 395–400.
11. Choi HN, Kim HH, Oh JS, *et al.* Factors influencing the miss rate of polyps in a tandem colonoscopy study. Korean J Gastroenterol 2014; 64(1): 24–30.
12. van Rijn JC, Reitsma JB, Stoker J, et a*l.* Polyp miss rate determined by tandem colonoscopy: A systematic review. Am J Gastroenterol 2006; 101(2): 343–350.
13. Bernal J, Sanchez J, Vilarino F. Towards automatic polyp detection with a polyp appearance model. Pattern Recognition 2012; 45(9): 3166–3182.
14. Younghak S, Balasingham I. Comparison of hand-craft feature based SVM and CNN based deep learning framework for automatic polyp classification. Annu Int Conf IEEE Eng Med Biol Soc 2017; 2017: 3277–3280.
15. Urban G, Tripathi P, Alkayali T, et al. Deep learning Localizes and identifies polyps in real yime with 96% accuracy in screening colonoscopy. Gastroenterology 2018; 155(4): 1069–1078.
16. Taha B, Werghi N, Dias J. Automatic polyp detection in endoscopy videos: A survey. Biomed Eng 2017.
17. Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2016; 2818–2826.
18. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. ICLR 2015; 1–14.
19. Zhang X, Zhang X, Li S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. p. 770–778.
20. Tajbakhsh N, Gurudu SR, Liang J. Automated polyp detection in colonoscopy videos using shape and context information. IEEE Trans Med Imaging 2016; 35(2): 630–644.
21. Bernal J, Sanchez FJ, Fernandez-Esparrach G, et al. WM-DOVA maps for accurate polyp highlighting in Colonoscopy: Validation vs. Saliency maps from physicians. Comput Med Imaging Graph 2015; 43: 99–111.
22. Silva J, Histace A, Romain O, et al. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer. Int J Comput Assist Radiol Surg. 2014; 9(2): 283–293.
23. Freedman JS, Harari DY, Bamji ND, et al. The detection of premalignant colon polyps during colonoscopy is stable throughout the workday. Gastrointest Endosc 2011; 73(6): 1197–1206.

Frontier Scientific Publishing Pte. Ltd.

Add: 21 Serangoon North Avenue 5,
#03-03 Ban Teck Han Building, Singapore (554864)
Email: jai@front-sci.com
Web: http://jai.front-sci.com/index.php/jai

ISSN 2630-5046
02
9 772630 504043