IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Interactive System For Product Purchase Through Gesture and Voice

1. Gaurav Zanpure, Computer department, P.E.S Modern College Of Engineering, Pune
2. Adesh Oak, Computer department, P.E.S Modern College Of Engineering, Pune
3. Bhavik Ransubhe, Computer department, P.E.S Modern College Of Engineering, Pune
4. Rithvik Poojary, Computer department, P.E.S Modern College Of Engineering, Pune

Guide : Dattatray Modani, Computer department, P.E.S Modern College Of Engineering, Pune

***Abstract* —** E-commerce websites are becoming extremely prevalent, with a significant volume of individuals embracing this medium. This seems to be due to the enhanced accessibility and simplicity of use that these networks provide when it start shopping for necessities as well as other things from the convenience of your own home. This has enabled many handicapped people and others with accessibility concerns to successfully remain at home and purchase different products that they require. This has resulted in a plethora of e-commerce companies fighting for consumers' focus and attempting to grow their customer base in order to deliver accurate alternatives with simplicity. The integration of speech and gesture-based searching for items on e-commerce platforms is one example of something like this. The paradigm of utilizing Convolution Neural Network for the recognition of the gestures is one of the most difficult approaches that require extensive computations along with the region of interest estimation. The obtained results are indicate better performance of the system through the use of Root mean square Error evaluation.

Keywords: *Convolutional Neural Network, Voice recognition and Gesture Recognition.*

I INTRODUCTION

In current history, the E-Commerce sector has experienced tremendous expansion in internet marketplaces. This may be due to advancements in the online medium, which have resulted in a huge rise in the subscriber base, rendering most Web application implementations extremely cost. This really is critical for the growth of the information infrastructure, which has focused on improving the usefulness and accessibility of the product's consumers. The huge reduction in the cost of internet-enabled gadgets has accelerated the expansion, resulting in a massive boom with in E-Commerce sector.

There has already been a great deal of study into how to develop this platforms much further and supply consumers with novel and efficient factors that characterize it apart from its contemporaries. As a result, there has been an abundance of research towards increasing the customer satisfaction and providing unique characteristics that can engage a bigger percentage of subscribers. This empowers the website to expand as well as provide additional services, making it far easier for customers to purchase from the convenience of their own residences.

Humans have been seeking to comprehend human physical communication for millennia and have been quite effective in revealing intricacies of evolutionary cognition. Computers as well as other technologies have been used to investigate human feelings and make generalizations regarding human behavior because of complicated psychological interaction and detailed modes of behavior. To get a better comprehension of human emotions, scientists are studying hand signals, gestures, and expressions.

Historically, retailing was much more of a commercial operation, with shop performance reliant on the achievement of the business engagements. Selling products, managing expenses, and reaching profitability were the main priorities. With the importance of client brand recognition rising, the strategic management focuses from merely marketing to creating an engaging consumer experience.

A customer who walks into a retail establishment does not just want to purchase something. He wants to go on a voyage of contact and participation with a brand throughout several procuring rostrums. Customers' perceptions of how a buying experience should then have been shaped in large part by new tech. From the accessibility among several information sources collecting to the capacity to explore various possibilities for acquisition, compare characteristics, and find excellent prices, reading assessment information assists the contemporary potential in a variety of ways. Is it possible to persuade customers even more by providing a visual and tactile experience?

Customers' buying habits have shifted dramatically in the last few years. Purchasing must be viewed as a brand contact and involvement narrative, according to the company. This necessitates the development of a large number of interaction points. These observations may be earned at various times throughout the purchase process. The greater the degree of contact and participation that a business can achieve, the stronger the brand narrative will be. Testimonials are real-life events that help people connect with a business on an unconscious level. After this link is established, the consumer will return to the establishment not just to shop but also to maintain the connection.

Gesture recognition is utilized in a variety of industries, including automobiles, smart appliances, home appliances, sport, and defense, smartphones, and sign language interpretation. Among the most interesting jobs in the computer vision technique is recognizing these motions, which is tough for the machine to do. This presentation can help with the development of a beneficial element in e-commerce platforms that allows for fast browsing and engagement while purchasing things.

This is paired with the use of voice recognition, which are collected and utilized to explore and interface with the goods by the system. The technique for speech recognition is exceedingly intricate and difficult to understand. Voice command identification is used in multiple human integrates directly. A diverse variety of voice aided or voice command implementations have been commercialized as a result of advancements in speech identification. Examples of such applications are motorized wheelchair control, voice controlled motors, voice activated appliances, and voice controlled smartphone apps. Many papers have appeared in the literature on voice command recognition which have been evaluated effectively to understand its implementation and the gesture based product navigation for achieving our approach.

This research paper segregates the section 2 for the evaluation of the past work in the form of literature survey. Section 3 as Proposed work, Results are discussed in Section 4 and this paper is concluded in Section 5.

II RELATED WORKS

D. Ryumin et al. [1] presented a design for a smart robotic trolley for supermarkets with a touchscreen and a multimodal user interface that includes sign language and acoustic voice recognition, as well as a software framework for collecting sign language databases using the Kinect 2.0 device. The gathered corpus TheRusLan contains recordings of 13 natural Russian sign language signers. Each signer performed the same 164 sentences five times. The corpus contains a total of 10660 samples. TheRusLan will be utilized for more research and tests on a Russian sign language identification system that will be unsegregated into the smart retail trolley's conversation system.

Text mining was used by Preeti Mehra in this research to provide a unified understanding of terms in the retail business, which has been using gesture control technologies. A two-step technique including the extraction of keywords by text mining and then clustering these keywords into clusters was used to analyze data from retail customer evaluations [2]. Retailers have used gesture control technology in a variety of ways, including Kinect for Windows Retail Shopping, 'Gesture Control Screen,' 'Virtual Fitting Rooms,' 'Retail Interactive Touch Screens,' 'Use of Visual Mirrors,' and 'Augmented Reality Window Display'. Customer happiness is directly influenced by the 'Mechanism of Gesture' Control. The findings revealed the keywords that influence customer happiness as well as the features of this technology that influence their perception and help them make purchasing decisions.

Based on collaboration with a community in the Okutama district of Tokyo, Japan, Y. Shimizu et al. presented an interactive information support system for the rental bicycle industry. The authors used two distinct types of robot partners: a concierge-type and a humanoid type with a tablet PC. Each robot companion has a distinct function to perform in conversation and suggestion [3]. First and foremost, the authors should address what robots should do and what people should do while providing customer support. Customers, for example, may find it difficult to respond to shop clerks' inquiries about their personal expenses for today. Customers can be asked by robot partners instead of shop assistants if shop staff do not wish to ask such a question.

D. Wu et al. introduced a unique Deep Dynamic Neural Network for continuous gesture detection on multimodal data that included image and depth data as well as skeletal characteristics In contrast to past state-of-the-art approaches, the authors do not rely on handmade features, which are time-consuming to develop, especially when done separately for each input modality [4]. Instead, deep neural networks are used to automatically extract useful information from data. Because the input data is multimodal, the presented model incorporates two separate feature learning methods: (1) Deep Belief Networks for skeletal feature processing and (2) 3D Convolutional Neural Networks for RGB-D data. In addition, the authors used an HMM to expand their feature learning model to include temporal relationships.

T. Du et al. suggested a deep learning framework for determining behavioral categorization. Deep learning offers strong feature extraction and categorization capabilities, and it has a high research value as a simulation of the biological neural network method framework. It is extremely important to recognize the human body motion, regardless of algorithm or application direction, using the advanced intelligence algorithm of deep learning and the information of human body movement acquired [5]. To perform gesture recognition, the RNN, LSTM, and GRU models are built. The experimental findings of the three models were compared. The results reveal that these approaches can identify hand motions in real-time, especially complicated gestures.

H. Long et al. introduce a deep learning-based gesture recognition framework that predicts the square anchor of the gesture and crop the sub-image from the original gesture using a two-stage recognition framework. The sub-image is recognized by the classifier network, and the category of the related gesture is determined [6]. The square anchor design is utilized to overcome the anchor deformation problem during scaling. Pruning and weight quantification techniques are used to further compress the model. The model approach may be applied in the embedded terminal while assuring the accuracy of gesture detection.

A computer vision technique is suggested by H. A. Jalab for operating a media player utilizing a neural network that identifies four hand gestures: play, stop, forward, and reverse. After capturing a frame from the webcam camera, skin segmentation in LAB color space was utilized to separate skin areas from background pixels. A fresh picture was produced that included the user's hand border. The form features of a hand gesture are described using a convex hull and corner detection [7]. A supervised back-propagation multi-layer feed-forward neural network was also utilized to classify user hand motions. The categorization was completed without the use of any special instruments, such as gloves or a marker. However, despite minor classification failures, the suggested system performed effectively in classifying the four users' hand gesture commands.

Faster R-CNN is used by H. Ruan et al. to recognize gestures based on pictures rebuilt with programmable metasurface. Depending on the gesture placements, the authors enhance the imaging of these gesture areas and achieve high accuracy recognition of 10 types of gestures using CNN and high-resolution pictures. Gesture detection and identification may be quite useful in assisting people's communication in various scenarios [8].

S. Hussain et al. suggest a transfer learning and vision-based hand gesture recognition technique for unidirectional dynamic motions, the approach was made more robust by omitting skin color segmentation, blob detection, skin area cropping, and centroid extraction. The pre-trained model in the presented work is VGG16, a CNN architecture. It has 13 convolution layers followed by three completely linked layers. A convolutional neural network (CNN) is a sort of feed-forward neural network whose connection structure is inspired by the arrangement of the animal visual brain [9]. Because the authors need to distinguish eleven different hand forms, CNN is trained as a classifier using the transfer learning approach. With an accuracy of 93.09 percent, the prototype was successfully tested on seven different volunteers in various backdrops and lighting situations.

The convolutional neural network framework in deep learning was utilized by Y. Gu et al. [10] to create an online teaching gesture identification model that can be used to recognize five distinct types of gestures that teachers will employ in online teaching settings. The model's recognition efficiency and accuracy are quite good. The model may be used to investigate the impact and quality of gestures in teaching, as well as the enhancement of educational robots' performance in human-computer interaction. However, at this time, the model is unable to distinguish between symbolic and metaphorical gestures that must be identified depending on the language utilized by the teachers or the actual teaching environment.

Y. Zhang provides a conceptual model to investigate the impact of traditional and virtual communities on customers' choice to purchase a product online. It makes theoretical as well as practical contributions [11]. It integrates conventional and virtual communities and presents a conceptual model to examine the aspects that affect a buyer's need recognition in online buying more thoroughly by evaluating the structure of virtual communities and their influence on FNR. In terms of practical contribution, it suggests that the firm should consider word of mouth in both the traditional and virtual communities. Both communities are critical for potential customers to identify their requirements and create buy intents.

X. Hu. et al. employ a large data analysis approach, the deep forest algorithm, to develop a prediction model using real consumer online buying behavior data, to predict customer purchase behavior in the context of online purchasing [12]. The outcomes of the online purchasing behavior prediction model depending on the deep forest algorithm are superior to other models in some circumstances, according to actual evidence. Because multiple models may be cascaded with deep forests, such as substituting the cascaded forest model with a linear regression model, the model in this research has the potential to enhance the classification prediction impact even further.

H. Cheng presents a research model that depends on social cognition theory to investigate the links between trust, perceived website complexity, and online buying behavior. Furthermore, earlier research has found that the online shop atmosphere is a major component affecting online purchases. However, perceived website complexity is given less weight. Therefore, this research considers perceived website complexity to be an environmental element and investigates the association between perceived website complexity and online purchasing behavior [13].

III PROPOSED METHODOLOGY

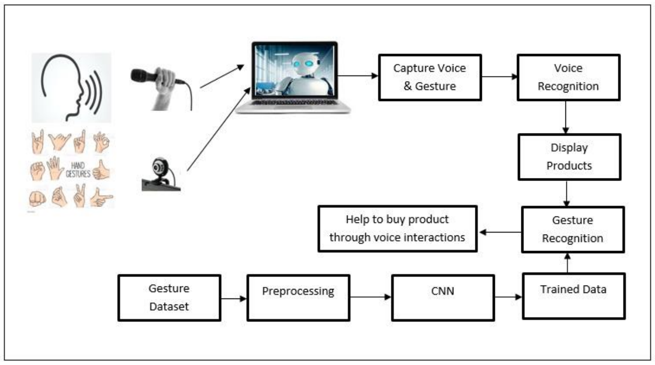


Figure 1: System Overview

The suggested approach for gesture-based shopping using hand gesture identification is detailed in the following step-by-step process. The methodology's system overview is depicted in Figure 1.

Step 1: Preprocessing – This is the initial phase in the process, in which the hand gesture images are taken using the OpenCv package. The cv2 library's VideoCapture capability is used to take photos of the specific hand gesture. There are five main forms of hand gestures: up, down, left, right, and buy. The YCbCr color model is used to identify the skin of the hand, and that particular region is clipped out. After that, the grayscale transformation is applied to the clipped hand gesture image. This produces a grayscale cropped picture, which is then downsized to 48x48 pixels and saved in a folder dedicated to that particular gesture. This process is performed iteratively for each of the 5 gestures for the purpose of achieving the input dataset.

Step 2: Image Segmentation – The input dataset, which comprises of acquired images that must be used for training, is made available in the previous stage. The training generator and the validation generator are in charge of this task. The training generator starts with a target picture size of 48x48 pixels, a batch size of 64, and a color mode of grayscale with categorical class mode.

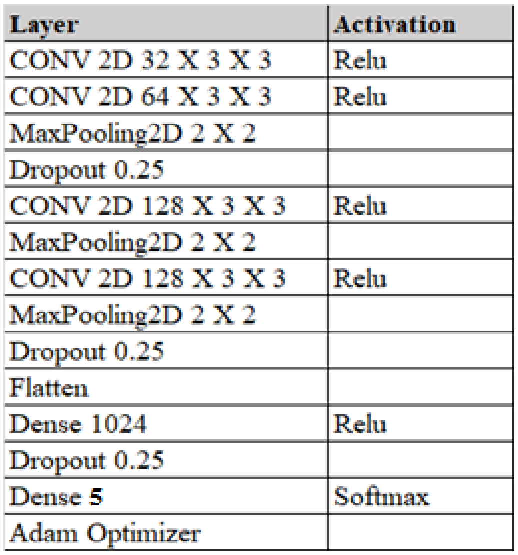
The validation generator is also created with comparable qualities, such as a target picture size of 48x48 pixels, a batch size of 64, and a color model of grayscale with a categorical class mode.

Step 3: Convolution Neural Network – This is the most important part of the suggested method, since it is in charge of detecting and identifying hand gestures. In the convolutional Neural Network module, the original image is utilized as an input. This model is trained using the input photographs that were acquired, preprocessed, and segregated in the previous rounds of the technique.

The input dataset is made up of the training and testing picture folders. Each folder is then separated into separate directories for each hand gesture and its associated photos. As component of the training phase, these photos are input into the CNN model. The photos should first be resized to 48 × 48 pixels in width and height. The model is trained on these images for 500 epochs with a batch size of 64 and a dense of 5 because we are considering 5 gestures in our presentation. The TensorFlow and Keras libraries are used in the python environment to facilitate the different components of the CNN model. The architecture may be seen in diagram 2 below.

After that, the CNN model created with this architecture is run for 500 epochs to produce a trained model file with the extension.h5.

Step 4: Decision Making – After achieving the trained model through CNN, the technique may now be evaluated for hand gesture detection for shopping. The camera is activated with the help of the OpenCV platform in order to crop the hand gestures. This cropped mage is transformed to grayscale and scaled to work with the.h5 file's contents. This procedure produces the matching hand gesture, which is then sorted with all of the matched gestures at the time. If the acquired count exceeds a certain threshold, the gesture is considered to have been detected and used for shopping purposes.



Layer	Activation
CONV 2D 32 X 3 X 3	Relu
CONV 2D 64 X 3 X 3	Relu
MaxPooling2D 2 X 2	
Dropout 0.25	
CONV 2D 128 X 3 X 3	Relu
MaxPooling2D 2 X 2	
CONV 2D 128 X 3 X 3	Relu
MaxPooling2D 2 X 2	
Dropout 0.25	
Flatten	
Dense 1024	Relu
Dropout 0.25	
Dense 5	Softmax
Adam Optimizer	

Figure 2: CNN network Architecture

IV. RESULTS AND DISCUSSIONS

The proposed methodology for achieving gesture based system through hand gesture recognition has been deployed on both the python programming language using the Spyder IDE and Java programming language using the NetBeans IDE. The approach utilizes the OpenCV, TensorFlow, and Keras, libraries to achieve the desired goals. The presented technique has been deployed on a computer consisting of 8 GB of RAM, 1 TB of Storage powered by an Intel Core i5 as the CPU.

For the evaluation purposes the proposed model is trained for 500 epochs on 5 gestures and the outcomes for the same need to be evaluated for the purpose of achieving effective performance of the approach.

The accuracy of the hand gesture recognition module needs to be assessed in order to determine the performance of the proposed approach. The accuracy of the approach can be calculated as a metric of error, as lower the error, greater is the accuracy. The evaluation of error can be effectively performed using RMSE metric.

The performance metric of RMSE or Root Mean Square Error is one of the most effective performance metric to determine the error achieved between a set of continuous and correlated attributes. The attributes being selected for the evaluation of the proposed methodology are, hand gesture identified correctly and hand gesture identified incorrectly. The RMSE is calculated using the equation 1 given below.

$$\mathrm{RMSE}_{fo} = [\sum_{i=1}^{N} (z_{f_i} - z_{o_i})^2 / N]^{1/2}$$

Where,

Where,

$\sum$ - Summation

$(Zfi - Zoi)^2$ - Differences Squared for the hand gesture identified correctly and hand gesture identified incorrectly

N - Number of conducted Experiments.

The RMSE values are computed for a number of iterations of hand gesture recognition performed through this proposed approach. Each of the 5 hand gestures are tested for the recognition 10 times. Each of the times the recognition output of the proposed approach is recorded. The outcomes are then utilized for the purpose of RMSE evaluation. These values of RMSE are rigorously calculated with the outcomes stipulated in the table 1 given below.

Gesture	Number of Iterations	Correctly identified hand Gesture	Incorrectly identified hand Gesture	MSE
Top	10	9	1	1
Bottom	10	8	2	4
Left	10	9	1	1
Right	10	7	3	9
Bye	10	10	0	0

Table 1: RMSE outcomes for 5 hand gesture recognition.

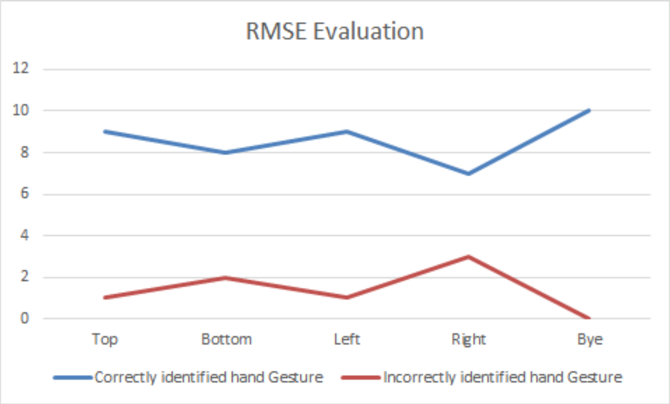


Figure 2: Line Graph for RMSE outcomes for 5 hand gesture recognition.

The results attained for the recognition performance and RMSE values in the table 1 given above are being utilized for the purpose of achieving the line graph in the figure 1 given above. The graph and table above illustrate the intended methodology for hand gesture recognition achieving an incredibly low error rate. The higher recognition accuracy may be attributed to the suggested approach's use of deep learning via CNN, which enhances recognition accuracy substantially. The hand gesture recognition error has an RMSE of 3.8, which is a really good outcome in the very first try of the proposed work.

III CONCLUSION AND FUTURE SCOPE

With the rapid development of computer technology, efficient human-computer interaction has now become an indispensable part of people's daily life. The most commonly used human-computer interaction mode is to rely on simple mechanical devices, such as mouse, keyboard, touch screen, etc. Through these typical controllers, it is difficult to achieve an immersive control experience. For example, handles or data gloves are usually used for interaction in a virtual reality environment, and bulky devices seriously affect the user's immersive experience. Therefore, the development of a more immersive human-computer interaction method has received widespread attention, and it is worth noting that the human computer interaction method through gesture recognition has been recognized by people. The hand is a complex deformable body with multiple degrees of freedom. The addition of voice based interaction also improves the overall user experience by improving the effective interaction between the user and the product. Hence, Convolution neural network is used to handle gesture and multi- threaded environment allows handling the graphical user interface efficiently.

In the future this model can be enhanced to work as the API and it can be enhance to deploy in real time shopping malls.

REFERENCES

[1] D. Ryumin, D. Ivanko, A. Axyonov, I. Kagirov, A. Karpov and M. Zelezny, "Human-Robot Interaction with Smart Shopping Trolley Using Sign Language: Data Collection," 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), 2019, pp. 949-954, DOI: 10.1109/PERCOMW.2019.8730886.

[2] Preeti Mehra and Balpreet Kaur, "Gesture Recognition: towards Making Future Retail Buying Experience Stimulating," International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8 Issue-6, March 2020.

[3] Y. Shimizu, S. Yoshida, J. Shimazaki, and N. Kubota, "An interactive support system for activating shopping streets using robot partners in informationally structured space," 2013 IEEE Workshop on Advanced Robotics and its Social Impacts, 2013, pp. 70-75, DOI: 10.1109/ARSO.2013.6705508.

[4] D. Wu et al., "Deep Dynamic Neural Networks for Multimodal Gesture Segmentation and Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 8, pp. 1583-1597, 1 Aug. 2016, DOI: 10.1109/TPAMI.2016.2537340.

[5] T. Du, X. Ren, and H. Li, "Gesture recognition method based on deep learning," 2018 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC), 2018, pp. 782-787, DOI: 10.1109/YAC.2018.8406477.

[6] H. Long, M. Liu, and M. Li, "Real-time gesture recognition in the complex background based on deep learning," 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 2020, pp. 2137-2140, DOI: 10.1109/ITAIC49862.2020.9339060.

[7] H. A. Jalab and H. K. Omer, "Human-computer interface using hand gesture recognition based on neural network," 2015 5th National Symposium on Information Technology: Towards New Smart World (NSITNSW), 2015, pp. 1-6, DOI: 10.1109/NSITNSW.2015.7176391.

[8] H. Ruan, M. Wei, H. Zhao, H. Li, and L. Li, "Programmable Metasurface Based Microwave Gesture Detection and

Recognition Using Deep Learning," 2020 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO), 2020, pp. 1-4, DOI: 10.1109/NEMO49486.2020.9343444.

[9] S. Hussain, R. Saxena, X. Han, J. A. Khan, and H. Shin, "Hand gesture recognition using deep learning," 2017 International SoC Design Conference (ISOCC), 2017, pp. 48-49, DOI: 10.1109/ISOCC.2017.8368821.

[10] Y. Gu, J. Hu, Y. Zhou, and L. Lu, "Online Teaching Gestures Recognition Model Based on Deep Learning," 2020 International Conference on Networking and Network Applications (NaNA), 2020, pp. 410-416, DOI: 10.1109/NaNA51271.2020.00076.

[11] Y. Zhang and Y. Feng, "Factors that Influence a Buyer's Decision Process of Shopping Online: The Effects of Tradition and Virtual Community," 2011 International Conference of Information Technology, Computer Engineering and Management Sciences, 2011, pp. 294-297, DOI: 10.1109/ICM.2011.316.

[12] X. Hu, Y. Yang, L. Chen and S. Zhu, "Research on a Prediction Model of Online Shopping Behavior Based on Deep Forest Algorithm," 2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD), 2020, pp. 137-141, DOI: 10.1109/ICAIBD49809.2020.9137436.

[13] H. Cheng and T. Fu, "The Determinants of Online Shopping Behavior," 2018 International Conference on Intelligent Autonomous Systems (ICoIAS), 2018, pp. 97-100, DOI: 10.1109/ICoIAS.2018.8494098.

*****