International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 2, Mar-Apr 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Spitting and Littering Detection System through Deep Learning

Mr.Vasim Tamboli[1], Vikhe Bhakti[2], Shirsath Ishwari[3], Kakade Shruti[4]

[*1]Asst.prof. Department of Computer Technology, Padmashri Dr. Vitthalrao Vikhe Patil Institute of Technology & Engineering (Polytechnic), Loni, Maharashtra, India

[*2,3,4]Department of Computer Technology, Padmashri Dr. Vitthalrao Vikhe Patil Institute of Technology & Engineering (Polytechnic), Loni, Maharashtra, India

--************************----------------------------------

Abstract:

All levels of government are currently working around the clock to contain the Coronavirus. Consequently, the system's primary objective is to employ cutting-edge computer technology to curb the spread of infectious diseases such as the Coronavirus by enforcing a severe ban on public spitting through the use of real-time drone-based surveillance systems. Spitting is the primary vector for the virus's maximum replication. A combination of closed-circuit television and unmanned aerial vehicles (UAVs) will monitor public spaces and record any actions taking place there in real time. Following transmission over the Cloud, the data will be evaluated for the presence of public spitting. The local authorities will have access to the results through a web/mobile interface. The person can thereafter face punishment from the authorities. A more effective management and monitoring of public spitting can be achieved with the help of this approach. Computer vision, namely Convolutional Neural Networks and Decision Making, has been employed in the system's design to accomplish the task of spitting and garbage littering detection.

Keywords: Image Normalization, Convolutional Neural Networks, Decision Making.

--************************----------------------------------

I. INTRODUCTION

The COVID-19 pandemic is the greatest threat to humankind right now. Every human life has ground to a halt. All throughout the world, millions of people are impacted. Despite this, there are individuals who refuse to accept the gravity of the situation and who persist in engaging in risky behaviors that put other people in harm's way. Spatting in public is one example of this kind of behavior.

Spitting is widespread in India and can transmit diseases such as COVID-19, the common cold, influenza, tuberculosis, hepatitis, and many more. Public spitting is now illegal in India according to the Disaster Management Act, but it will become much more common once the lockdown limitations are relaxed, leading to a dramatic increase in the number of COVID-19 cases in the country. In order to contain a pandemic, the suggested solution employs cutting-edge technology to restrict this improper behavior.

People carelessly spit out things like tobacco, saliva, and water in public spaces. This kind of thing releases dangerous bacteria and viruses into the air. The existing method of dealing with this problem is laborious and wasteful. In densely populated places, such as train stations and bus terminals, law enforcement must apprehend the perpetrators. People argue with the authorities rather than pay any fines. Unlawfully fining innocent persons is another common practice by the authorities.

To overcome the shortcomings of the existing procedures, this method suggests implementing CCTV/drone surveillance systems in public places such as bus terminals, train stations, etc., to identify and penalize those found endorsing such practices.

[1] The study by CUN CHEN et al. on scaling behavior and war outbreak prediction sheds light on the underlying mechanisms of human society. The author uses statistical tools and nonlinear analysis to turn history into a predictive and analytical science. This will help us understand social systems better and build better societies.

[2] The authors BENEDETTA PICANO et al. postulated a learning-based paradigm to simulate human perception by taking into account all five senses and the multimodal nature of human cognition. The conditional probability distribution of having a perception level for each sense, assuming particular individual attributes, has been identified using the mixed density network. For sensory integration to take place, the neurocognitive literature suggests a multi-sensory perception paradigm. After that, the users were grouped according to their combined perceptual sensitivity using a supervised learning module. Finally, the suggested framework has been applied to the HVR services in order to demonstrate the benefits of a brain-aware decision-making scheme compared to the more typical brain-agnostic method. This case study is an attempt to demonstrate how performance can be enhanced by taking system reliability into account in order to ensure a perceptual target deadline is met.

[3] Human motion pattern recognition using multi-features and inertial measurement unit (IMU) was the primary emphasis of FAHIMA HAJJEJ et al. To begin, we gather information from two datasets—WC and REALDISP—that have been carefully chosen. The next step is to preprocess the raw data using a state-of-the-art IMU filtering technique, which involves removing minimum and average gravity. After that, the data that was filtered has been divided up using various sorts of windows, and the suggested model will use a 5-second window. In order to identify active and passive motion patterns, the segmented data is then fed into the dynamic time warping algorithm. In addition, for active pattern features, we have chosen spectral roll off and SST, whereas for passive pattern features, we have chosen TECC and spectral flux. The features vector reduction process also makes use of an OFNDA optimization method. Last but not least, LSTM has been used to optimize features based on both publicly available datasets. The model achieved exceptional results, with mean recognition accuracies of 84.68% for active patterns and 87.85% for passive patterns. With an average accuracy of 86.26 percent, the suggested model is well-suited for usage in intelligent healthcare education resources. The suggested model improves human motion recognition when compared to traditional state-of-the-art approaches. Future Aims and Constraints: Nevertheless, this research isn't without its flaws; for example, it can't account for all possible environments, there are delays in responses, and it can't identify all motion patterns. In the future, the author hopes to incorporate a wider range of motion patterns into our model through the use of various types of sensors. This includes smart settings, healthcare services, and outdoor complexes. In order to enhance the results, the author will conduct additional experiments using other domain features and patterns identification methodologies. Not to mention the significant computational cost and maximum time it takes for the system to respond due to the proposed filter. The author plans to use several methods for gyroscope-based rotational angles to enhance the filter even further.

The study is structured into five parts. We present a comprehensive overview of ITS in the second section. A concise introduction to deep learning and its uses is given in Section 3. Section 4 details how smart cities and ITS make use of deep learning to detect pedestrians, summarizes the work of several academics in the area, and lays out the obstacles to further study in each subfield. Next, in Section 5, we shall present the final verdict.

II. LITERATURE SURVEY

[4] Creating a reliable HAR system for practical use was the primary focus of the study by JAEYEONG RYU et al. the author's primary focus was on action data and training data. The author began by investigating how different types of training data affected recognition. Our DBC dataset was

compared with the EBR dataset by the author. The outcomes demonstrated the need for application-specific details in the training dataset. Two, the author considered the data type's stability as it relates to motion variation data. In order to test the data of an untrained individual using the cross-subject methodology, the author created a virtual environment. The angle data was useful for gathering broad patterns of movement based on the outcomes of the tests. Therefore, in the cross-subject procedure, the angle-based recognition system attained greater accuracy. Analyzing the effect of training data and data kinds in real-world applications is the core contribution of our work. Authors wishing to employ the HAR system should ensure that their training data contains examples of nuanced behaviors and that their action data conveys examples of more generalized patterns. The author employed precise action data to rule out the impact of the sensing environment. The author plans to incorporate lower-quality action data gathered from a camera-based sensing environment in a future study, which will expand the scope and limitations of this work. The reliability of the angle measurements will suffer as a result. The performance of our recognizer will be improved by the author using a data augmentation strategy. With all the different kinds of motion variations included in the augmented motion data, the recognition system can be made more robust.

[5] YAN CHEN et al. suggests a way to recognize human behavior that relies on an improved attention mechanism. We suggest a better attention module after studying its drawbacks in relation to the current channel attention method. Experiments are conducted from several angles, such as visualization results, network accuracy improvement, new network parameters, etc., to confirm that the enhanced attention module is successful. The functionality of cross-structure learning is confirmed by employing a multi-scale convolution kernel to extract behavior characteristics across various receptive fields. Subsequently, the convolution, pool, and full connection layers are rationally designed to further refine these characteristics. The impact of model structure on the impact of soft migration is examined, and the necessity of a multi-stage progressive supervision method is confirmed by comparing the supervision at different phases. If the structure of the monitoring network is comparable to that of the learning network, then the network is more likely to converge. The scope and limitations of future work include the use of additional sensors to increase the data dimension and, by extension, the recognition accuracy. Future work will concentrate on ways to make the model lighter, as it has many parameters in our method's model module.

[6] The sixth source is Xin Liu and colleagues. Proteomic investigations frequently employ PSSM and other features following basic processing (vectorization from 2 to 1). Few studies have focused on how to effectively extract features from PSSM, though. The author of this work suggested an approach called RF-PSSM. It involved using 2DPCA to extract useful features from PSSM, and then a prediction model was built using rotation forest. The RF-PSSM demonstrated an adequate level of prediction ability in the experiments. against further validate 2DPCA-PSSM's effectiveness, the author compared it against PSSM and four other feature representation approaches. To top it all off, the author compared RF-PSSM to SVM and other state-of-the-art methods, and RF-PSSM came out on top. Lastly, the author utilized RF-PSSM to identify novel proteins that could interact with E1. These findings could inform future wet research. Visit https://github.com to see the dataset and source code. URL: com/flyinsky6/RF-PSSM. The obvious flaw in this research is that it does not provide any specific software or websites to implement the technique. Future Work: The author plans to use graph neural network technology to improve the prediction of HCV-host interaction in several areas, including protein structure, physicochemical properties, and the author's focus on improving the performance of predictions in the field of HCV-host interaction.

[7] In their proposal for online person identification, ZAMAN WAHID et al. suggest

using human micro-expression, a new social behavioral biometric. In order to derive probability ratings for six different emotion classes from microblogs written by OSN users, an emotion detection model is created. We create and apply a new way to describe emotion signals in order to capture human microexpressions as they evolve over time. The suggested strategy is evaluated by training three classical ML algorithms following emotion-progression feature extraction. The last step is to enhance person identification performance using a rank-level fusion technique that is based on weighted majority voting. With a rank-1 accuracy of 61.73%, the experimental results show that human micro-expression, the suggested social behavioral biometric, has a strong and distinctive ability for online person identification. Furthermore, the suggested approach outperformed the original state-of-the-art SBB qualities in terms of performance. To see if the suggested human micro-expression feature could improve overall identification performance, the author is currently merging it with features like URL, retweet, friendship network, and others. Further improvement in recognition performance may be possible via fine-tuning transformer-based models for improved emotion extraction, which may lead to improved person identification performance. It is also possible to conduct additional research into how the suggested method performs when using other tweet batches to generate emotion signals. To further enhance the efficacy of person identification, it may be worthwhile to investigate high-dimensional emotion-progressions.

[8] A lens-less imaging system is suggested by THUONG NGUYEN CANH et al. as a means to safeguard visual privacy from perceived attacks by the human visual system. Measurements with similarity loss and/or patterns with TV, inheritability, and RIP losses allowed the author to achieve this goal. The suggested methods successfully strike a balance between human-imperceptibility and recognition accuracy for lens-less imaging, as shown by experimental results from both simulation and the hardware prototype as well as subjective study. We, the authors, think we have built the first ever learnable hardware-based recognition system to take visual privacy protection into account. This report presents initial findings that highlight difficulties in merging robust privacy protection with excellent identification accuracy. Future Work: (1) utilizing a bigger dataset, (2) creating more effective hardware implementations, and (3) exchanging visual privacy for more accuracy could enhance recognition accuracy in future work.

[9] In order to recognize human actions, USMAN AZMAT et al. The segmentation and feature extraction method developed by Felzenszwalb forms the basis of the proposed system. The author took a novel method in our feature extraction module by using GMM-EM-based elliptical modeling to target each affected body part independently. In order to achieve outstanding algorithmic outcomes, the suggested method makes use of deep learning and machine learning. Not only does the suggested system work well for action identification, but it also has additional applications such as stance estimation and body part segmentation. Future Aims/Restrictions: We intend to investigate more intricate human action recognition scenarios and investigate additional features for multihuman-based systems in the near future. Furthermore, author ai intends to use deep learning techniques to enhance the efficiency of labeling.

[10] A sensory wristwatch-like device was created by HONG-QUAN NGUYEN et al., who also set a new standard for hand gesture detection using a camera on the wrist. The enormous number of participants, wide variation of gestures within classes, high levels of resemblance across classes, and variety of indoor and outdoor settings make this dataset difficult to work with. Some have speculated that it was the initial dataset used to train and test deep learning models. The author also assessed various convolutional neural network (CNN) models for hand gesture identification that rely on vision. These models range from the most traditional 3D ones, like C3D and R3D, to more modern, resource-efficient ones, like MobileNet3D,

EfficientNet3D, and MoviNets, as well as lightweight ones, such R3D-18 and R(2+1) D18. According to our results, the MoviNet variation achieves very good accuracy. The Top-1 accuracy increased from 1.36% to 3.37% using our suggested solution, which integrates optical flow and RGB in a two-stream MoviNet. As a result, several SOTA activities in the domain of video action classification could benefit from this method's use to improve their efficiency. Designers can use this information to their advantage when deciding which model to deploy on edge devices. The author plans to create an app and test it with subjects in various settings and lighting conditions in the future. Before contacting the recognizer, the author additionally studies strategies to swiftly detect the movements.

[11] All authentication systems aim to accurately determine a person's identity, according to BAHA A. ALSAIFY et al. (2011). This paper presents a novel approach to person-aware identification detection that makes use of representative features extracted from recorded Wi-Fi signals affected by the motions of the subject. Finding the activity with the best identifying accuracy was the goal of this investigation, which included walking, standing, and sitting on a chair, as well as picking up a pen from the ground. According to the results of the trials, whether you're sitting, standing, or walking, you'll get very high levels of accuracy, with standing and sitting yielding marginally superior results. When the standing and sitting on a chair activity was done, the accuracy rate was 100%, compared to 99% when the walking exercise was done and 64.5% when the pen was picked up off the ground.
Future Plans: The author plans to build upon the work in this publication by exploring human identification in Non-Line-of-Sight (NLoS) circumstances. This will increase the system's reliability and expand its range. The impact of the subject's location on the system's overall reliability is another area that the authors intend to examine. In particular, the author plans to answer the following question: What effect does the subject's closeness to the transmitter or receiver have on the system's overall identification accuracy? Along with answering the first question, the author also intends to find out whether the acquired dataset can be used to extract data on acceleration and velocity. Can accurate models of human identity be constructed using measured velocities and accelerations? We place a premium on researching the feasibility of alternative methods.

[12] A novel method for human identification has been proposed by NOVA EKA BUDIYANTA et al., which involves processing raw 3D LiDAR point cloud data before running it through a single-stage 2D detector on projected cylindrical coordinates along the vertical axis. The shape of the 3D LiDAR point cloud—a cylindrical plane with the LiDAR device as its center axis—is relied upon by the projection technique. Based on criteria such model size, loss, accuracy, and inference speed, the best model was selected in the present study using YOLOv5, a one-stage 2D object identification framework. Different YOLOv5x versions of the YOLOv5 series have different model sizes; the smallest of these is the YOLOv5n. Overall, the model loss is acceptable, being near to 0, according to the training and testing results. In terms of accuracy, the YOLOv5x model is the most accurate with a mAP of 79.83% and an inference speed of 25 ms. On the other hand, the YOLOv5n model is the least accurate with a mAP of 44.35% and an inference speed of 7.6 ms. Moreover, YOLOv5s outperforms YOLOv5n in terms of performance, with a mAP of 58.54% and an inference speed of 8.3 ms. Finally, the YOLOv5l model has a mAP that is approximately 3 percentage points lower than that of YOLOv5x. To be more precise, it does 76.88% mAP and 18 ms inference speed, while the YOLOv5m model does 71.66% mAP and 12.2 ms inference speed.

[13] To improve the accuracy of quantifying human mistake, ZHAO Changxiao et al. used Bayesian networks, evidential reasoning, and a fuzzy multiple attributive group decision-making approach. Nonetheless, STPA-CREAM differs from the operation chain in that it explains the HCI process using a hierarchical security control

framework. This keeps the model from becoming oversimplified and brings it closer to the real scene. Concurrently, the safety control framework allows for the thorough consideration of the impact of multi-party interaction on HRA and the identification of additional cause scenarios.

[14] The 3DCNLSTM neural network, developed by CHHAYA GUPTA et al., combines multiple methods. Additional methods are also examined in the study. A number of datasets have been utilized to compare different methods with the suggested model; these include UCF101, NTU-RGB-D, KITTI, and NTU-RGB-D 120. Additionally, a fused dataset was created using data fusion techniques. This dataset has 79282 photos from 148 classes and 184680 videos from 281 classes. Additionally, the suggested model is contrasted with others that have been trained on the same datasets: DTR-HAR, CNN with BiLSTM, Classic CNN, and YOLOv6. In terms of accuracy, the proposed model is superior to all previous models. The suggested model is a unified neural network that integrates four distinct methods: data fusion, feature extraction, object detection, and skeletal articulation. The recommended model has been enhanced through the application of multiplicative LSTM. Limitations and Potential Future Work: Deeper convolutional networks could improve feature extraction in the future, and this model could be used in conjunction with humanoids. An elderly person living alone in their house can benefit from this model since it can monitor their daily activities.

[15] MISHA KARIM et al. examined HAR from multiple angles, such as user behavior, lifestyles, and static and dynamic living activities. There has been little focus on real-time HAR for healthcare security, though. Problems with developing HAR systems include data scarcity, technical constraints, and the complexity of real-time activities. It takes a lot of processing power to keep up with the demands of reacting to changing surroundings and public or populated locations in real time. Current research also fails to address real-time activities, and there is a lack of high-quality real-time data.

[16] According to Shubham gade et al. the trained model is obtained by deploying an LSTM model in the encoded data; the temporal fusion transformer (TFT) makes efficient use of it. Target value retrieval via TFT is accomplished by use of a look-back window of a fixed duration. Then, using permutations, TFT forms time-dependent properties of the external input and builds the unknown input. Finally, TFT provides contextual metadata about the things under evaluation that is independent of time by using static variables. The developed model is able to provide accurate predictions because of this integration. This integration of LSTM model with the fusion of CNN can enhance the performance of the deep learning models efficiently.

METHODOLOGY

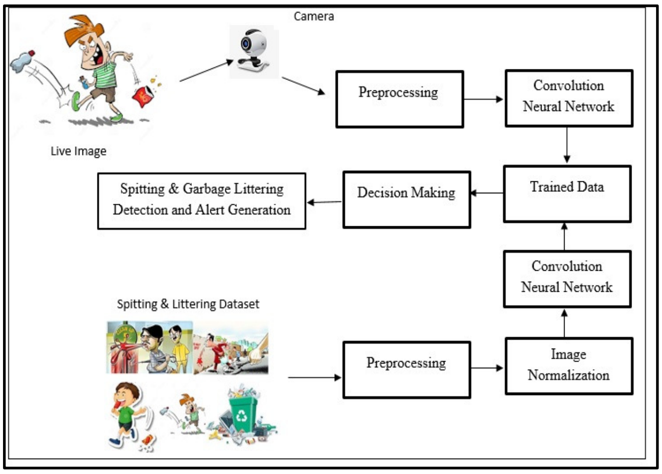

Figure 1: Proposed method block diagram

The method that has been suggested to establish a Spitting and Littering detection system depicted in the system overview in Figure 1 up top. The suggested method was based in part on the execution of the procedures detailed below.

Step 1: Dataset preparation: As a first step in developing the system, this involves collecting data on various body language expressions, including spitting, littering, normal behavior, and blank posture. This dataset will be used for testing and training purposes. In order to train the Convolution neural network, the acquired images are first downsized to 48 by 48 dimensions and then turned into grayscale.

Step 2: Image Normalization: The photographs are downsized to 48 X 48 dimensions before the training begins.

At this stage of the model's suggested implementation, we construct an Image Data Generator object using the keras and tensorflow libraries with a ratio of 1/255 to do the detailed image analysis. The training and testing of Gesture/posture images both involve this technique. Parameters such as the paths to the training and testing directories, the dimensions of the image, a batch size of 64, the color mode of "grayscale," and the class mode of "Categorical" are used to segment the Image Data Generator object.

Step 3: *Training with Convolution Neural network* : The tensorflow library's Sequential class is used to set up a model for a sequential neural network. After that, for each picture dimension, the first layer of the neural network is a convolution layer with a 32-kernel, 3-by-3-inch model using the "Relu" activation function. Afterwards, a second convolution layer is incorporated, this time with 64 3 X 3 kernels and the "Relu" activation function. An additional 2 x 2 max pooling layer is included, with a 25% dropout percentage.

Additionally, a max pooling layer of size 2 X 2 and a third layer of convolution with 128 3 X 3 kernels using the "Relu" activation function are incorporated. At last, a 4th Convolution layer is incorporated, which consists of 128 3 X 3 kernels activated by the "Relu" function. The next step is to incorporate a 2 X 2 max pooling layer with a 25% dropout ratio.

Lastly, a dense layer of size 1024 with the "Relu" activation function is used to stop the neural network using the flatten function. After training the convolution neural network, we use a "softmax" activation function and a 2 Dense layer to set a 50 dropout percentage.

We employ the Adam optimizer with 1000 epochs for gesture/posture images to improve the result during training. The model makes use of the H5 file, which stores the trained data after training, during testing. The Neural Network's architecture is illustrated in Figure 2.

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 4	Softmax
Adam Optimizer	

Figure 2: Convolutional Neural Network Architecture

Step 4: Testing through Decision Making: With the help of live streaming from the openCV library in Python, the data from the trained model is read into the testing image in h5 file format during testing. A model of this test image is created as an object for the neural network. The classes are predicted using this data in an integer format. Postures such as spitting, littering, being blank, and normal are identified in the four-class lexicon. Then, using the Python Pywhatkit Library, the picture of the offending individual will be sent to the admin's Whatsapp number if spitting or littering is discovered.

III. RESULTS AND DISCUSSIONS

The proposed method for Spitting and Littering Detection System through Deep Learning.was developed using the Anaconda framework, Python, and the Spyder IDE. The development computer has 1 terabyte of secondary memory and 8 gigabytes of main RAM. A number of factors have been considered in order to determine how feasible the proposed plan is. In this part, we detail the results of the experimental study.

The obtained results for Accuracy and loss are depicted below in the following figures.

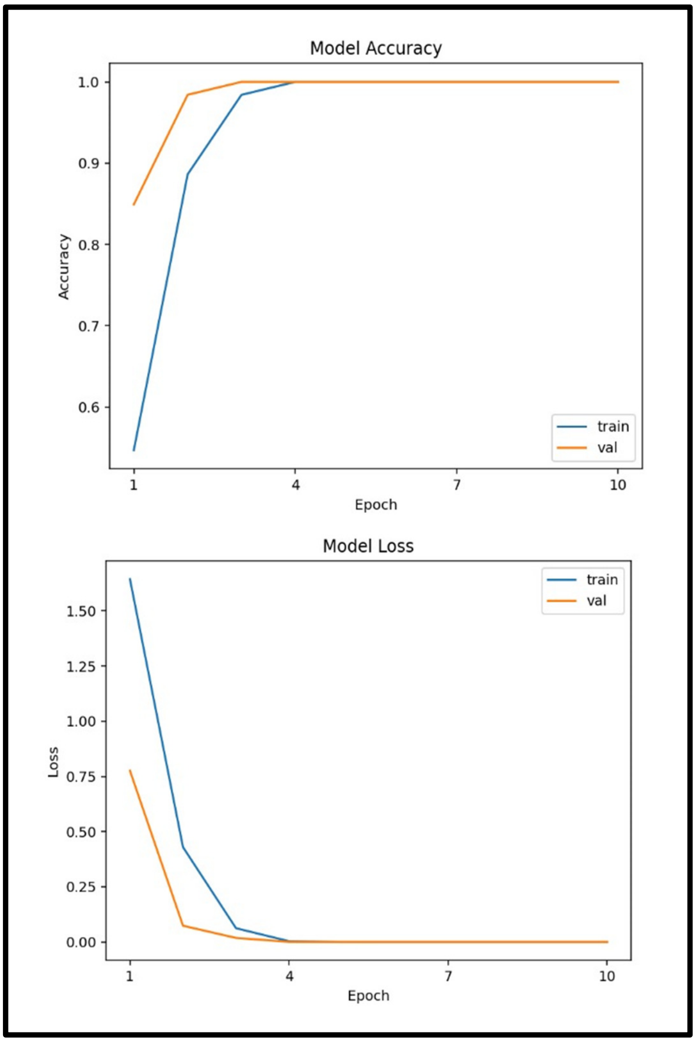

Figure 2: Accuracy and Loss

IV. CONCLUSION AND FUTURE SCOPE

About 1,49,886 people have died in India as a result of the COVID-19 epidemic, and 1.86 million people have died worldwide. Unless the virus is able to be contained, this number will continue to rise. In order to keep an eye on public spaces and notify the proper authorities in a fast and predictive manner if someone is spotted spitting, the system makes use of numerous modern techniques, including the Internet of Things, cloud computing, big data, and artificial intelligence. As a result, people will be less likely to litter in public places, which will improve air quality and health. This is why we've developed a method that uses Convolutional Neural Networks and Decision Making in the realm of image processing and computer vision to effectively detect spitting and littering. The following are some ways this project could be expanded in the future to accommodate new areas of research: API for convenience in integrating and evaluating In the future, this method can be applied in real-life situations, such as near roads, temples, hospitals, and so on.

For the purpose of future research directions this Project can be extends with the below mentioned points

- ✓ API for easier integration and evaluation
- ✓ For the future implementation this technique can be used in the real time scenario like on the road, near temples, hospitals etc.

REFERENCES

1] C. Chen, N. N. Xiong, X. Guo and J. Ren, "The System Identification and Prediction of the Social Earthquakes Burst in Human Society," in *IEEE Access*, vol. 8, pp. 103848-103859, 2020, doi: 10.1109/ACCESS.2020.2999575.

2] B. Picano, "Multi-Sensorial Human Perceptual Experience Model Identifier for Haptics Virtual RealityServicesinTactfulNetworking,"in *IEEEAccess*,vol9,pp.147549147558,2021,doi:10.1109/ACCESS.2021.3124607.

3] F. Hajjej *et al.*, "Deep Human Motion Detection and Multi-Features Analysis for Smart Healthcare Learning Tools," in *IEEE Access*, vol. 10, pp. 116527-116539, 2022, doi:10.1109/ACCESS.2022.3214986.

4] J. Ryu, A. K. Patil, B. Chakravarthi, A. Balasubramanyam, S. Park and Y. Chai, "Angular Features-Based Human Action Recognition System for a Real Application With Subtle Unit Actions," in *IEEE Access*, vol. 10, pp. 9645-9657, 2022, doi: 10.1109/ACCESS.2022.3144456.

5] Y. Chen, "Human Behavior Recognition Based on Multiscale Convolutional Neural Network," in *IEEE Access*, vol. 11, pp. 13533-13544, 2023, doi: 10.1109/ACCESS.2022.3209816.

6] X. Liu, Y. Lu, L. Wang, W. Geng, X. Shi and X. Zhang, "RF-PSSM: A Combination of Rotation Forest Algorithm and Position-Specific Scoring Matrix for Improved Prediction of Protein-Protein Interactions Between Hepatitis C Virus and Human," in *Big Data Mining and Analytics*, vol. 6, no. 1, pp. 21-31, March 2023, doi: 10.26599/BDMA.2022.9020031.

7] Z. Wahid, A. H. Bari, F. Anzum and M. L. Gavrilova, "Human Micro-Expression: A Novel Social Behavioral Biometric for Person Identification," in *IEEE Access*, vol. 11, pp. 57481-57493, 2023, doi: 10.1109/ACCESS.2023.3283932.

8] T. N. Canh, T. T. Ngo and H. Nagahara, "Human-Imperceptible Identification With Learnable Lensless Imaging," in *IEEE Access*, vol. 11, pp. 95724-95733, 2023, doi: 10.1109/ACCESS.2023.3308069.

9] U. Azmat *et al.*, "An Elliptical Modeling Supported System for Human Action Deep Recognition Over Aerial Surveillance," in *IEEE Access*, vol. 11, pp. 75671-75685, 2023, doi: 10.1109/ACCESS.2023.3266774.

10] H. -Q. Nguyen *et al.*, "Hand Gesture Recognition From Wrist-Worn Camera for Human–Machine Interaction," in *IEEE Access*, vol. 11, pp. 53262-53274, 2023, doi: 10.1109/ACCESS.2023.3279845.

11] B. A. Alsaify, M. M. Almazari, R. Alazrai, O. Y. Al-Jarrah and W. Al Zibideh, "Wi-Fi Signals for Passive Human Identification: A Study of Three Activities," in *IEEE Access*, vol. 12, pp. 113087-113098, 2024, doi: 10.1109/ACCESS.2024.3443231.

12] N. Eka Budiyanta, E. Mulyanto Yuniarno, T. Usagawa and M. H. Purnomo, "Utilizing a Single-Stage 2D Detector in 3D LiDAR Point Cloud With Vertical Cylindrical Coordinate Projection for Human Identification," in *IEEE Access*, vol. 12, pp. 72672-72687, 2024, doi: 10.1109/ACCESS.2024.3402227.

13] C. Zhao, H. Li, W. Zhang, J. Dai and L. Dong, "Risk Identification and Safety Assessment of Human-Computer Interaction in Integrated Avionics Based on STAMP," in *Journal of Systems Engineering and Electronics*, vol. 35, no. 3, pp. 689-706, June 2024, doi: 10.23919/JSEE.2024.000031.

14] C. Gupta *et al.*, "A Real-Time 3-Dimensional Object Detection Based Human Action Recognition Model," in *IEEE Open Journal of the Computer Society*, vol. 5, pp. 14-26, 2024, doi: 10.1109/OJCS.2023.3334528.

15] MISHA KARIM, SHAH KHALID, ALIYA ALERYANI, JAWAD KHAN, IRFAN ULLAH, AND ZAFAR ALI, “Human Action Recognition Systems: A Review of the Trends and State-of-the-Art”, Received 12 February 2024, accepted 26 February 2024, date of publication 7 March 2024, date of current version 13 March 2024. Digital Object Identifier 10.1109/ACCESS.2024.3373199.

[16] “Enriching Crop Yield and Price Prediction Using Transformers,” Shubham Gade, Amita Singh, Sanjay Patil in International Journal of Research in Engineering, Science and Management, vol. 7, no. 12, pp. 87-93, December 2024.

*****