Computer Vision And Deep Learning Based Hand-Hygiene Guidance

Rajvardhan Sudhakar Deshmukh
School of Computer Science and Engineering
Vellore Institute of Technology, Vellore, Tamil Nadu, India
Email- rajvardhan.sudhakar2017@vitstudent.ac.in

Rishav Agarwal
School of Computer Science and Engineering
Vellore Institute of Technology, Vellore, Tamil Nadu, India
Email- rishav.agarwal2017@vitstudent.ac.in

Syed Moaz Murtuza
School of Computer Science and Engineering
Vellore Institute of Technology, Vellore, Tamil Nadu, India
Email- syedmoaz.murtuza2017@vitstudent.ac.in

Preetha Evangeline D.
School of Computer Science and Engineering
Vellore Institute of Technology, Vellore, Tamil Nadu, India
Email- preetha.evangeline@vit.ac.in

Abstract- In the year 2020, the world was plagued with a fatal virus that made the people helpless in every way. A simple touch scared everyone and a seemingly harmless cough made everyone run a mile away. A cure seemed like a distant thing but precautions were the only way forward. Various health organizations started providing precautious steps in order to avoid the affliction of the coronavirus. One of the simpler ones, as provided by the World Health Organization (WHO), is the right way to sanitize and wash our hands. Something as simple as this could go a long way in saving lives and it is the one thing that is most overlooked. Hand Hygiene has been termed as one of the most effective ways to curb infections outside and inside hospitals. It is often not done accurately, thus, yielding no positive results. The project aims in creating a product which can be installed in various public and private places which detects whether any individual is following the proper hand hygiene 6-steps as prescribed by the WHO. The technology of Computer Vision and Deep Learning is used to bring the project to fruition. An open-source dataset of 32471 annotated videos of hand washing steps is used in order to train the neural network models. To test and validate the steps, Computer Vision libraries are used which detect the region of interests, segments it and then tests it in the trained model. This will validate each of the 6 steps and the time frame in which it is carried out before going on to the next step. Various Convolutional Neural Network models are employed in order to realize the best approach to getting the most accuracy for the same. This Hand Hygiene Guidance System not only ensures that proper washing of hands is carried out but also encourages people to maintain such hygiene as frequently as possible.

Keywords – Hand Hygiene, Infection, Video Capture, Computer Vision, Deep Neural Networks

I. INTRODUCTION

The most common act of medical prevention against any infectious disease is the act of hand washing and hand hygiene. The spread of many diseases can be easily prevented simply by washing one's hand and taking care of its hygiene. Even though it is the most effective in disease prevention, it is often neglected as no one realizes the benefit of this act [1]. There are countless ways that germs can infect the hands. The real danger of any illness prevails when those germs get transmitted into our body via our mouths, eyes as was the case during the COVID-19 pandemic, or infects the food and water that we intake. In case of medical personnel or patients, hospital acquired infections are the most common types of infections [2]. The more negligence towards hand hygiene, the riskier it is for the health of the person. If the pandemic of 2020 has taught the world anything it is that simply by ensuring hand hygiene, a catastrophe can be prevented.

The COVID-19 pandemic made the world realize the importance of hand hygiene throughout the lifetime. Humans often neglect the importance of hand hygiene unless they are compelled to do so. The main problem that needs to be solved is how accurately the act of hand washing is being carried out and for how long. The World Health Organization has prescribed a 6-step hand washing technique which ensures prevention of infections [3]. There is a need to develop a Hand-Hygiene Guidance system based on a Deep Learning and Computer Vision approach using the 6-step approach as the baseline for detection. This approach should not only validate the act of hand washing but also instill in the minds of people the importance of the same, thus, increasing their frequency of carrying out a hand hygiene process. The work aims at creating an integrated system which will promote hand hygiene while also ensuring that the act of hand washing is being done accurately and efficiently. A small act of ensuring hand hygiene will go a long way in keeping the world safe from a disastrous pandemic. The work presented aims at developing a Hand-Hygiene Guidance system based on a Deep Learning and Computer Vision approach. The project has the following objectives which have been completed during the fulfillment of the work: To create a Deep Neural Network (DNN) architecture which will be trained and tested using a dataset of 32471 annotated videos for the purpose of classifying the steps performed by a user during hand wash, to capture the steps in the act of hand-washing as a live feed using the concepts of Computer Vision (CV) and also guide the user through the 6 steps prescribed by the World Health Organization (WHO) for a hygienic hand wash, to validate whether the steps undertaken are accurate and thresholds are met and to integrate the DNN and CV modules to create a system that promotes hand hygiene and ensures that the task is done accurately.

II. Related Works

The work of S. Yeung et. al. [4] talks about developing a system which keeps a check on the hand hygiene compliance in a Hospital environment. They propose a Machine Learning algorithm approach using Convolutional Neural Network and Computer Vision. The paper deals with depth perception and detects if a person is performing the act of hand washing in front of a soap dispenser. The positive and negative result for the same is based on their posture and how much are they in the frame while performing the task. This removes the need for keeping a manual registry of compliances, thus, saving valuable time of the staffs. The drawback of this paper is that it is unable to detect how accurate is the hand hygiene steps or how long is the person actually performing the task. According to the work of M. R. Islam et. al. [5] Deep Convolutional Neural Network is a prominent approach in the aspect of hand gesture recognition. It talks about human computer interaction and the need for feature recognition. It deals with the recognition of the American Sign Language using Multi Class Support Vector Machines providing with an accuracy of 94.57%. The work of S. Khan et. al. [6] gives a comprehensive guide on the usage of Convolutional Neural Networks (CNN) for Computer Vision. It deals with various aspects of the CNN models as they are applied in the concepts of Computer Vision such as image classification and object detection. A. Ullah et. al [7] proposed a novel activity acknowledgment technique by handling the video information utilizing convolutional neural networks furthermore, a deep bidirectional LSTM network. The proposed technique is fit for learning long-term groupings and can handle extensive recordings by examining highlights for a specific time span. The work of M. Zhang et. al. [8] has utilized computer vision to acquaint a mechanized methodology with video investigation of careful execution. A convolutional neural network architecture for object discovery was utilized to identify working hands in open medical procedure recordings. According to W. Luo et. al. [9] a novel deep neural network can mutually reason about 3D identification, tracking, and movement determining given information caught by a 3D sensor. Their methodology performs 3D convolutions across existence over an elevated perspective portrayal of the 3D world, which is exceptionally effective as far as both memory and calculation. In the work of K. Roy [10] a two-stage deep learning based the methodology is introduced to identify hands in unconstrained situations. To additional upgrade the yield of the hand identifier they proposed a convolutional neural network (CNN) based skin identification methodology which decreases the events of false positives.

The paper by J. Cheng et. al. [11] proposes a quick and precise video object division technique that can instantly begin the division cycle once receiving the pictures. A likeness-based scoring capacity is embraced to refine these article parts by contrasting them with the visual data in the principal outline. In the work of M. Izadpanahkakhk et. al. [12] a novel methodology is proposed where convolutional neural networks (CNNs) alongside move learning are used. The modules are a pre-prepared CNN engineering as a component extractor and an AI classifier. The investigations exhibited that the ROI extraction module could fundamentally track down the fitting palm print ROIs, and the check results were extremely accurate. The work of K. Gu et. al. [13] presents the first study of picture model strength to the minuscule changes found across video outlines. They also found that a greater part of the fragility found in recordings lies outside the commonplace definition of antagonistic models (99.9%). In the paper by C. Chen et. al. [14], they introduced deep computing of extremely low light crude videos: around one lux of illuminance. To supplement this line of work, they gathered another dataset of raw dark videos, in which high-resolution raw information is caught at a video rate. Via cautiously planning a learning-based pipeline and presenting another loss function to improve temporal stability, they prepared a Siamese network on static raw video with the end goal that the network sums up to video of fluid scenes at test time. The work of A. Rössler et. al. [15]

presented a novel face manipulation dataset of about a large portion of 1,000,000 altered pictures (from over 1000 recordings). Utilizing the new dataset, they presented benchmarks for traditional picture legal assignments, including grouping and division, considering recordings compacted at different quality levels. The paper by A. Khamis [16] introduces a radio-based device-free system (RFWash) in order to monitor the hand hygiene technique in accordance with the standard compliance by the World Health Organization (WHO). A deep model is utilized which is then trained in naturally occurring hand hygiene gestures. According to the paper by A. Haque et al. [17] a vision based non- intrusive technique is the approach towards a smart hospital. It aims at reducing hospital acquired infections by eliminating close-proximity detection of compliances. They work with depth images in order to differentiate between a person performing the hand hygiene act and a person just there in the vicinity. They ensure that the privacy is maintained as the image received does not reveal the environment or the color of the surroundings. The paper showcases promising results in case of accounting for an increase in hand hygiene via the proposed method. The work by M. Kim et. al. [18] talks about the usage of 3D Convolutional Neural Network coupled with Computer Vision in order to facilitate hand hygiene in hospitals. They detected the region-of-interest of an anesthesiologist's hand rubbing technique and used it to create the classifiers. The moderated an accuracy of 76% using a simple classification technique.

III. Proposed Algorithm

3.1 Related Concepts –

The application for determination of the quality of hand wash based on the 6 steps prescribed by the World Health Organization (WHO) requires processing of video, classifying its scenes using Deep Neural Networks and a software for logically integrating these separate parts into one.

Convolutional Neural Networks (CNN): CNN is a Methodology which consumes input, and then segregates the different aspects within the input depending upon the user's discretion. Its architecture was derived from the electrical impulses between neurons within the Human Brain [19]. VGG16 has 13 convolutional layers and 5 pooling layers, where the pooling layers are utilized to compress the image. This method has been instrumental for classification accuracy and very high performance [20]. MobileNetV2 model is a CNN which provides a very efficient network architecture which can be adjusted to meet the demands of mobile vision applications which require low number of parameters to provide good accuracy [21]. ResNet is a trained CNN which uses a concept known as 'short-cut' in building blocks. Due to which it outperforms conventional deep CNNs which are generally known for their high testing errors [22].

Computer Vision: In this, image extraction takes place when feature detection techniques try to identify features based on the intensity patterns in the input image. And depending on the user's requirement, some of the quality for which these patterns are to be chosen can be valued over the other such as Repeatability and Quantity [23]. Thresholding Contours which is a method of converting a grayscale or RBG image into an image which is binary, which is done by setting a threshold within an image and if the pixel exceeds the threshold it coverts that pixel into black or white [24]. In OpenCV, both of these tasks are inbuilt within its library which the user can use to implement them on any given image. It has more than 2500 algorithms which have been optimized and also include a plethora of modern machine learning and machine algorithms [25].

3.2 Architecture for the proposed system –

The application for determination of the quality of hand wash based on the 6 steps prescribed by the World Health Organization (WHO) requires processing of video, classifying its scenes using Deep Neural Networks and a software for logically integrating these separate parts into one.
First step is to detect when to call the API for classifying the scene i.e., determining the step which is being performed. To do so, we trigger our application (the set of actions for hand hygiene detection) when there is any kind of movement in front of the live feed taken from the camera mounted above the wash basin. Once the movement is detected we narrow the region of interest (ROI), capture the frame and send it to the trained model for classification. The Deep Neural Networks (DNN) is an integral part of the application which is utilized to create a model for the above-mentioned classification (i.e., step is being performed by the user) in a repetitive fashion due to the dynamic nature of the application. The custom model created will be trained on an open dataset created from 32471 videos (annotated properly with codes for the steps prescribed by the WHO) which have a frame rate of 30 fps. Frames along with their annotation are extracted from each of the videos and then stored in a folder which will then be divided into a train, test and validation set for the sake of training the custom model. After the classification of the current scene, we repeat the process until the decided threshold time for the classified step is crossed. Once it is done, the user is notified that this particular step is completed on the display screen in front of him/her and is directed to follow the rest of the steps in an orderly fashion.

The processing of all the live feed and changes on the displays will be done with the help of OpenCV, and the task involving neural network will be completed with the help of Tensorflow as it provides with inbuilt, robust and easy to use commands for model creation and training [26]. Along with the custom model for the classification of scenes we will also utilize a few standard neural networks like the AlexNet, MobileNetV2 and Xception in order to provide a comparative analysis of the results (i.e., based on accuracies, nature of the neural network and the training time) [27].
Furthermore, the above-mentioned flow of the application provides extreme flexibility, scalability and robustness. In case, there are any changes in the amount of threshold time then there is no need for any changes to the model. Only the logic in the software needs to be changed according to the problem statement's requirements. Also, the custom model created will work for any additional steps prescribed by WHO (which is unlikely due to the high-level of scrutiny that the WHO guidelines go through) provided a proper dataset is being used for training.

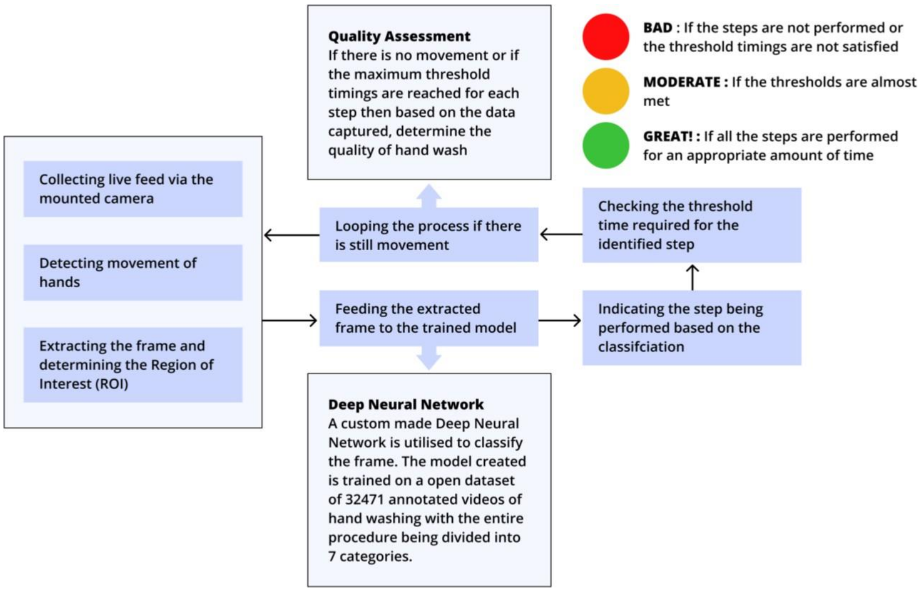


Figure 1. State Diagram of the proposed model

IV. Proposed System And Analysis Design

4.1 Introduction –

It is often seen that monitoring and assurance goes a long way in conditioning a positive feedback loop in humans [28]. If a particular act is monitored and then validated then humans get conditioned into repeating the same activity in a more enthusiastic manner. A system which guides a person into the steps to wash hand and validates each step once it is done will ensure that each individual takes the act seriously [29]. In order to do the same, the work proposes a technique where the action of washing hands is captured as a live video stream and then each step is validated by a marker. The steps that are proposed for washing hands are the 6 steps prescribed by the World Health Organization (WHO). When manipulation of video streams or capturing and detecting the same comes into play, then the concept of Computer Vision is always required. The open-source library OpenCV is the most prominent tools which provide all kinds of facilities in order to work with video streams. The work will use OpenCV in order to detect the hand movement and capture the same which will be used ahead for validation process. Furthermore, the concept of Neural Networks will be utilized by the work in order to create models based on pre-existing dataset (open-source dataset of 32471 videos) which will be used to validate the steps with utmost accuracy.

4.2 Product Features –

The product is based on Neural Network and Computer Vision algorithms. It has features that are defined by the administrators' perspectives and also from the users' perspectives.
The following are the features of the product created:

- **Flexibility of dataset usage:** This feature of the product is pertinent to the developer or the administrator side. The product has been developed using the open source dataset (32471 videos) of hand washing technique following the WHO mandated steps. The product allows for addition of further datasets making it more efficient in its working.

- **Real-time video capturing and guidance:** This feature of the product is pertinent to the consumer side or the people performing the act of hand washing. The video of them performing each step is captured by the product using a webcam, which then further goes for testing to verify whether it has been performed for the accurate time and in an accurate manner. This is carried out using Computer Vision (OpenCV).
- **Validation:** Once the video is captured, the product verifies whether that particular step was accurately carried out. If it was, it validates the consumer or the user to move on to the next step in hand hygiene process. The validation is carried out using Neural Network algorithms.

4.3 User Characteristics –

The product is aimed towards the common people and the medical personnel in order to provide them and to increase awareness about proper hand hygiene mechanism especially during the times of the COVID-19 pandemic.
The different user classes and their characteristics are as follows:

- **Developers:** These users are defined by the developers of the product or the software. They have all the development tools installed in their system. They will be the ones training the data and creating a model to test the data. They will also be writing a code to capture the hand movements and extract the region of interest using OpenCV. Finally, they will also be responsible for installing the system at the appropriate locations.
- **Hospital Medical Personnel:** These users are the frontline workers who are actively participating in the hospitals in order to treat the incoming patients. Even without a global pandemic their well-being and hygiene is of utmost importance, thus, having a system like this will help them to consciously time themselves and follow the accurate steps as mandated by the WHO.
- **Individuals (Patients/ Non-patients):** These basically include every single individual in and around the area where the product is installed. Being a system which encourages hand hygiene in accurate manner, it is important for patients to ensure that they are performing that act accurately to prevent any infections. These users can also be individuals sitting at home or washing hands in public rest rooms.

4.4 Product Requirements –

- **Efficiency** The efficiency of the product can be categorized as Time Efficiency and Space Efficiency in the following manner:
  Time: As per the WHO guidelines, an average of 5 seconds is required for each of the six steps of hand hygiene, thus, the entire act of efficient hand hygiene takes about 30 seconds. This time frame is short and is efficient as it provides the users with the desired outcome.
  Space: The prototype version of this product would simply requires a display, camera and a small pre-programmed processing unit, thus, taking minimal space. In the sense of deployment with proper hardware, the product would be installed on top of a sink at eye level. This signifies that the space complexity of the entire product would be very less as it will be using unused space over a particular sink or any hand washing system in place.
- **Reliability:** As the testing of the Neural Networks models will be done, a comprehensive result will be produced which will state the most efficient and reliable model for the testing and training of the data. In the case of capturing the live video feed, the only problem that might arise is if more than one individual tries to enter frame. But all in all, the product is expected to be reliable in its use. On top of that, the steps are mandated by the WHO, thus, they have gone through enough scrutiny to be considered reliable.
- **Portability:** The product is intended to mount near or on top of a sink. As the usage of this product is intended to be for the duration of the entire day, the hardware portability is also high as it requires only a small display and camera and pre-programmable processing units like the Google AIY Vision Kit or the AWS DeepLens.
- **Usability:** The usability of the product is simple, as the level of expertise to operate the same is low. The technical aspect of it will be taken care of by the developers but the usage of it does not require skilled resources. The user just has to follow the WHO mandated steps being displayed and ensure that their act of washing hands is within the frame of the video.

4.5 Implementation Requirements –

The implementation methodology that will be followed for the product is the Parallel Implementation technique. The two modules of the product is coded separately then integrated with an interface by the use of PyFlow, to come up with the final product.

- **Neural Network Models:** In order to test and train the dataset of 32471, a system of high RAM is required. Even though a prototype can be created with RAM size of 16-32 GB, it is better if a platform like the Google Cloud Platform (GCP) is made available. Transfer Learning is applied on the three models- VGG16, ResNet and MobileNetV2- are used for the purpose and then compared to realize the best model to implement.
- **Computer Vision:** In order to get the input data, Computer Vision algorithms are required so as to capture the video content and segment the images. These images act as inputs for the model, which then verifies it via the pre-trained model and validates the WHO mandated steps for hand washing.

4.6 System Requirements –

The prototype has been created on an DELL Inspiron 7460, with an x64-based Intel®; CoreTM i5-7200U CPU including 16 GB RAM, a 64-bit Windows as well as Linux operating system and an NVIDIA 940MX graphics card. A webcam is of utmost importance for the capturing of the video image. The training of the classifier was performed on the GPU provided by Kaggle and Google Colab. To ensure that the system runs for more epochs and provides better accuracies, it is advised to work on a higher powered system which has higher RAM and more than one GPUs allowing for the extensive processing of the dataset. The entire product is coded in Python 3.9 which requires the following packages: tensorflow, OpenCV (cv2), numpy, sklearn, PyFlow. Depending on the computation power, the coding is done in varying notebooks like Google Colab or Keras Notebook.

V. Results And Discussion

The system is implemented with image processing functions from OpenCV. Transfer Learning is applied on three different CNN models for the training and testing of the dataset- the VGG16, ResNet and MobileNetV2 models, which are pre-trained on the open source dataset ImageNet. The open dataset was taken from a project, by the Ministry of Education and Science, Republic of Latvia, "Integration of reliable technologies for protection against Covid-19 in healthcare and high-risk areas." According to the dataset, column 1 signifies the video frame, column 2 signifies whether the person is washing or not (0/1) and the third column defines the step the user is in (1-6), as shown in Table 1.

Table 1: Dataset Description

Code	Code explanation
1	Hand washing movement – palm to palm
2	Hand washing movement – Palm over dorsum, fingers interlaced
3	Hand washing movement – Palm to palm, fingers interlaced
4	Hand washing movement – Backs of fingers to opposing palm, fingers interlocked
5	Hand washing movement – Rotational rubbing of the thumb
6	Hand washing movement- Fingertips to palm

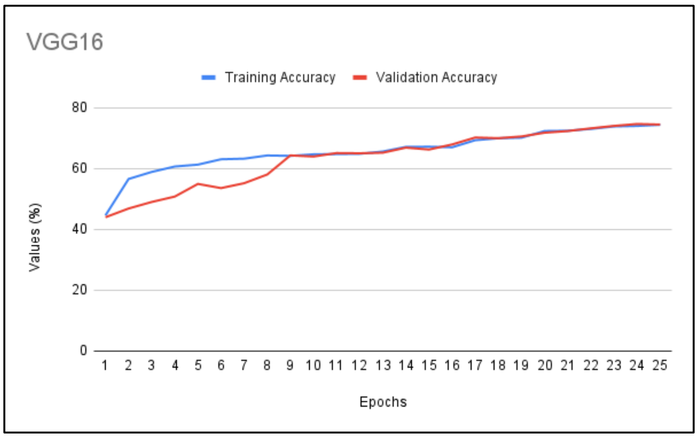
107	3500	1	3
108	3533.333	1	3
109	3566.667	1	3
110	3600	1	2
111	3633.333	1	2
112	3666.667	1	2

Currently, the created models are benchmarked with the performance evaluation metrics, namely the Training Accuracy and the Validation Accuracy. The three CNN models for the given dataset are evaluated using these metrics. The models are trained and tested using the same dataset, which has been adapted from the Kaggle platform [30]. Due to the unavailability of a better platform like the GCP or AWS, or the use of GPUs for a longer period of time, the testing of the models could be facilitated only for 10 Epochs. Table 2 shows the different values of the Accuracies across the three models at each of the 10 epochs.

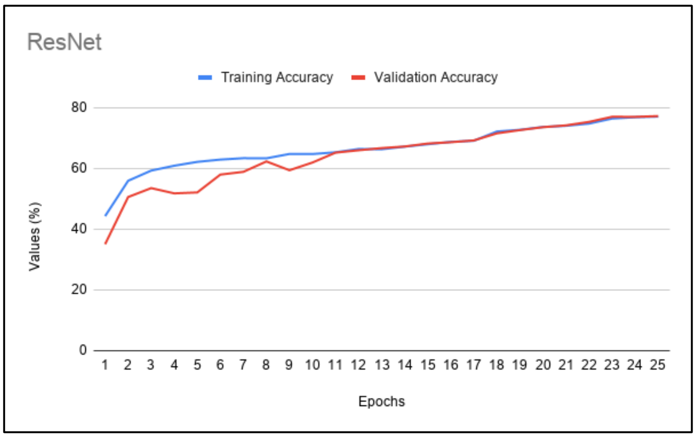
Table 2: Accuracies of the three models (in %)

Epochs	VGG16		ResNet		MobileNetV2	
	Training Accuracy	Validation Accuracy	Training Accuracy	Validation Accuracy	Training Accuracy	Validation Accuracy
1	44.61	44.00	44.31	35.07	57.77	50.26
2	56.58	46.87	55.95	50.60	60.15	55.42
3	58.90	49.02	59.30	53.53	61.68	53.31
4	60.68	50.80	60.90	51.80	62.42	50.10
5	61.30	54.96	62.17	52.11	63.12	51.73
6	63.06	53.58	62.92	57.98	63.91	58.82
7	63.26	55.19	63.38	58.89	64.64	58.24
8	64.31	58.05	63.34	62.31	64.55	61.72
9	64.18	64.33	64.77	59.41	65.58	65.08
10	64.68	63.96	64.75	61.96	65.48	68.58
11	64.77	65.10	65.32	65.21	65.66	65.58
12	64.86	65.03	66.41	65.99	65.96	65.79
13	65.61	65.22	66.32	66.70	66.32	66.10
14	67.13	66.88	67.18	67.23	66.73	66.90
15	67.18	66.26	67.97	68.20	67.87	66.69
16	67.01	67.93	68.75	68.65	67.70	67.39
17	69.33	70.21	69.13	69.22	67.91	67.48
18	69.97	69.99	72.16	71.56	68.19	68.28
19	70.15	70.56	72.73	72.63	68.26	68.41
20	72.34	71.79	73.66	73.59	69.37	68.93
21	72.44	72.34	74.04	74.18	69.80	69.77
22	73.03	73.23	74.77	75.33	70.03	69.81
23	73.85	74.03	76.43	77.01	69.96	70.31
24	74.02	74.66	76.86	76.98	70.46	70.15
25	74.39	74.48	77.07	77.24	70.81	70.69

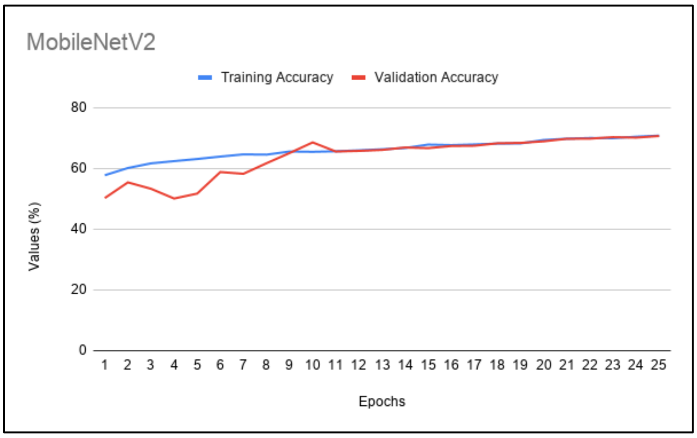
Based on Table 1, the following graphs are plotted which illustrates the comparison between the Training Accuracy and the Validation Accuracy for each of the 25 epochs and for each of the three models mentioned. Figure 2 (a) illustrates the comparison in case of VGG16 model. Figure 2 (b) illustrates the ResNet model and Figure 2 (c) illustrates the MobileNetV2 model of Convolutional Neural Network. After the said comparisons, Figure 2 (d) shows a final graph which compares the accuracy of each of the three models at 25 epochs.



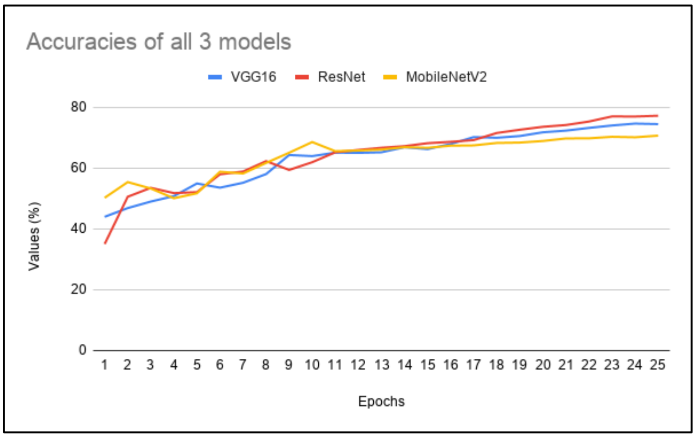
(a)



(b)



(c)



(d)

Figure 2. Visualization of the evaluation metric scores that are obtained by the models: (a) VGG16; (b) ResNet; (c) MobileNetV2; (d) Accuracy comparison of all the models

Figure 2(d) clearly shows that accuracy variation of each of the models. Initially, for only 10 epochs, the average accuracy of VGG16 was 60.16%, while ResNet shows an accuracy of 60.18%. Whereas, MobileNetV2 showcased the highest average accuracy, for 10 epochs, that of 62.93%. Thus, at that stage of the product it showed that MobileNetV2 has the best performance. But as the number of epochs was increased to 25, MobileNetV2 suggested a steady and a flat plot, as shown in figure 2(c), as compared to the other two models. Whereas, VGG16 and ResNet suggested an increasing trend in their curve with an increase in the number of epochs, as shown in figure 2(a) and figure 2(b). Thus, after 25 epochs the ResNet has the best result with the highest accuracy of 77.24%, followed by VGG16 with 74.48%, while MobileNetV2 had the least accuracy that of 70.69%.

Now, once the models are trained and tested and the best model realized, that is ResNet, the product implementation moves on to the Computer Vision aspect of it. Using OpenCV algorithms of thresholding and contouring along with background subtraction, the system will extract the video feed and break into frames which are preprocessed according to the input requirements of our trained classifier used to validate the steps prescribed by the WHO.

Figure 3 displays a block diagram which is a 1-dimensional visualization of the entire product. It displays how the video capturing system will be mounted over the sink, and the location of the screen where the validation of each steps would be highlighted.

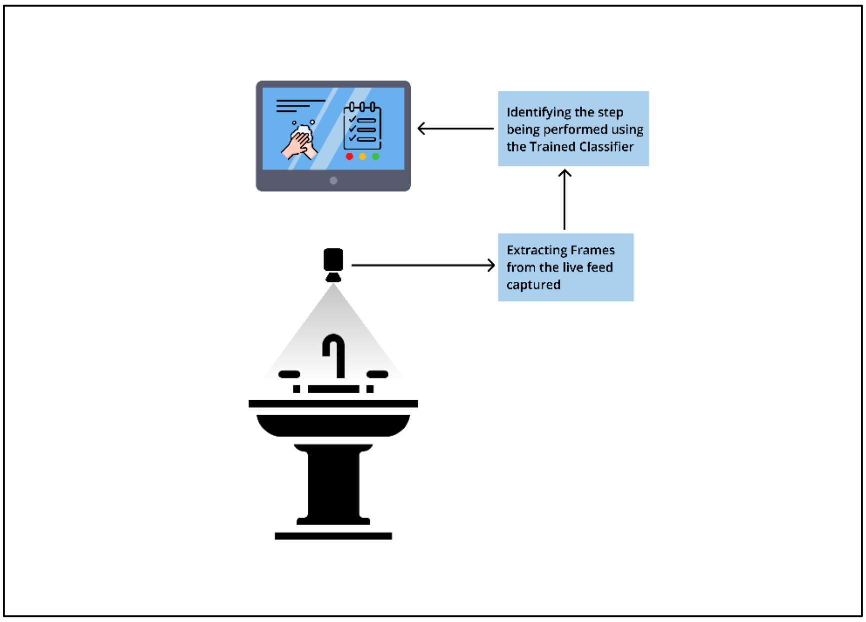


Figure 3. Visualization of the evaluation metric scores that are obtained by the models: (a) VGG16

Figure 4 displays the interface which has been created to capture the frames and also validate the user when they follow the steps correctly.

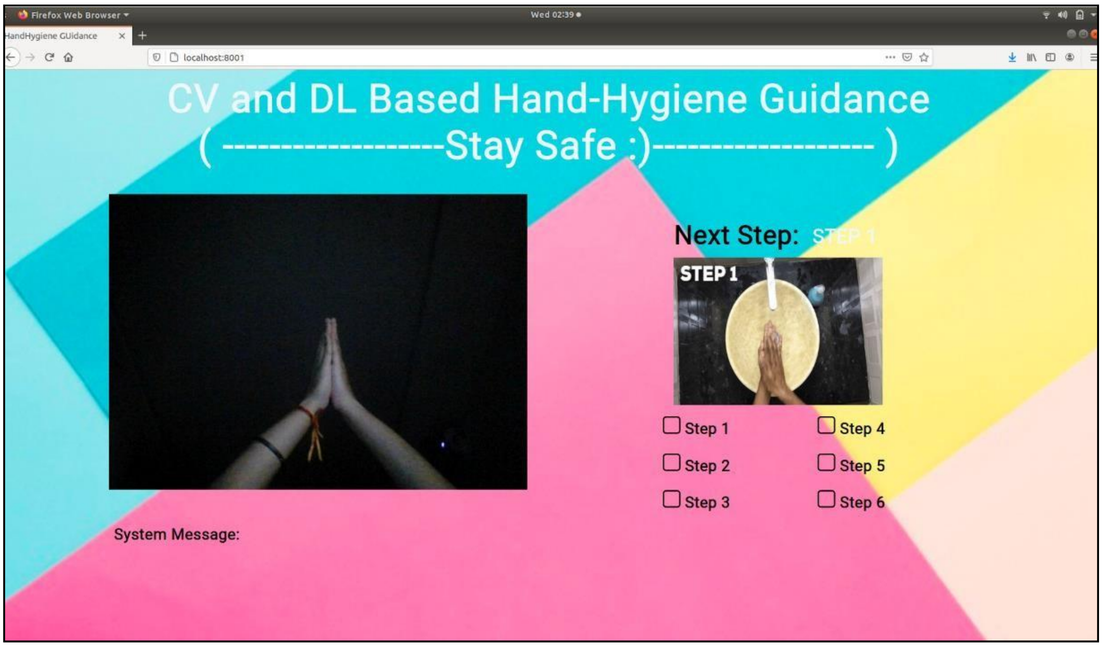


Figure 4: Interface for the capturing of hand movements

Once the above interface is integrated with the CNN model, a product is assembled, as shown in figure 3, which works in the following manner:

- The user puts their hands under the camera which is installed.
- They enact the step as displayed on the right side.
- As the threshold is met for enacting the step, the system validates the step and marks the step as completed in the checklist that is created.
- Once a particular step is validated, the user moves on to the next step, as the act of the next steps is displayed.

VI. CONCLUSION

The pandemic that swept the world into chaos, made everyone realize that one cannot compromise on their hygiene. According to the WHO, the guidelines provided by them for hand wash must be followed always and not just during some specific instance like the pandemic of 2020.

The implementation, done by using Deep Convolutional Neural Network and Computer Vision, provides a full proof system that not only guides the users through the process of a hygienic hand wash but also encourages them to do so with the help of the live interface that is being provided. This system also enforces a habit of carrying out the task efficiently thus conditioning them to do it without needing validation. The work realizes that ResNet is the best CNN model in order to train the dataset taken, that of 32471 videos. Using 25 epochs, it shows an accuracy of 77.24%, greater than the ones shown by VGG16 and MobileNetV2. It can also be taken into consideration that with better hardware, increase in GPUs supporting more number of epochs, the accuracy might subsequently increase, but it is expected that ResNet will still show the highest accuracy based on the trend observed.

The system that is created is easy to build and easy to use, requiring minimalistic robust hardware that can be found in any electronic store which makes it extremely robust and portable. Furthermore, due to the unique modularized structure of the software it becomes easy to make changes to the thresholds and the classified steps independently and with ease which makes the software flexible.

VII. REFERENCES

[1] Desai, A. N., & Patel, P. (2020). Stopping the spread of COVID-19. *Jama*, *323*(15), 1516-1516.

[2] Schuchat, A., Covid, C. D. C., & Team, R. (2020). Public health response to the initiation and spread of pandemic COVID-19 in the United States, February 24–April 21, 2020. *Morbidity and mortality weekly Report*, *69*(18), 551.

[3] Musu, M., Lai, A., Mereu, N. M., Galletta, M., Campagna, M., Tidore, M., & Coppola, R. C. (2017). Assessing hand hygiene compliance among healthcare workers in six Intensive Care Units. *Journal of preventive medicine and hygiene*, *58*(3), E231.

[4] Yeung, S., Alahi, A., Haque, A., Peng, B., Luo, Z., Singh, A., & Li, F. F. (2016). Vision-Based Hand Hygiene Monitoring in Hospitals. In *AMIA*.

[5] Islam, M. R., Mitu, U. K., Bhuiyan, R. A., & Shin, J. (2018, September). Hand gesture feature extraction using deep convolutional neural network for recognizing American sign language. In *2018 4th International Conference on Frontiers of Signal Processing (ICFSP)* (pp. 115-119). IEEE.

[6] Khan, S., Rahmani, H., Shah, S. A. A., & Bennamoun, M. (2018). A guide to convolutional neural networks for computer vision. *Synthesis Lectures on Computer Vision*, *8*(1), 1-207.

[7] Ullah, A., Ahmad, J., Muhammad, K., Sajjad, M., & Baik, S. W. (2017). Action recognition in video sequences using deep bi-directional LSTM with CNN features. *IEEE access*, *6*, 1155-1166.

[8] Zhang, M., Cheng, X., Copeland, D., Desai, A., Guan, M. Y., Brat, G. A., & Yeung, S. (2020). Using Computer Vision to Automate Hand Detection and Tracking of Surgeon Movements in Videos of Open Surgery. *arXiv preprint arXiv:2012.06948*.

[9] Luo, W., Yang, B., & Urtasun, R. (2018). Fast and furious: Real time end-to-end 3d detection, tracking and motion forecasting with a single convolutional net. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (pp. 3569-3577).

[10] Roy, K., Mohanty, A., & Sahay, R. R. (2017). Deep learning based hand detection in cluttered environment using skin segmentation. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 640-649).

[11] Cheng, J., Tsai, Y. H., Hung, W. C., Wang, S., & Yang, M. H. (2018). Fast and accurate online video object segmentation via tracking parts. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7415-7424).

[12] Izadpanahkakhk, M., Razavi, S. M., Taghipour-Gorjikolaie, M., Zahiri, S. H., & Uncini, A. (2018). Deep region of interest and feature extraction models for palmprint verification using convolutional neural networks transfer learning. *Applied Sciences*, *8*(7), 1210.

[13] Gu, K., Yang, B., Ngiam, J., Le, Q., & Shlens, J. (2019). Using videos to evaluate image model robustness. *arXiv preprint arXiv:1904.10076*.

[14] Chen, C., Chen, Q., Do, M. N., & Koltun, V. (2019). Seeing motion in the dark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3185-3194).

[15] Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2018). Faceforensics: A large-scale video dataset for forgery detection in human faces. *arXiv preprint arXiv:1803.09179*.

[16] Khamis, A., Kusy, B., Chou, C. T., McLaws, M. L., & Hu, W. (2020, November). RFWash: a weakly supervised tracking of hand hygiene technique. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems* (pp. 572-584).

[17] Haque, A., Guo, M., Alahi, A., Yeung, S., Luo, Z., Rege, A., & Fei-Fei, L. (2017, November). Towards vision-based smart hospitals: a system for tracking and monitoring hand hygiene compliance. In *Machine Learning for Healthcare Conference* (pp. 75-87). PMLR.

[18] Kim, M., Choi, J., & Kim, N. (2020). Fully Automated Hand Hygiene Monitoring\\in Operating Room using 3D Convolutional Neural Network. *arXiv preprint arXiv:2003.09087*.

[19] Tu, Z., Xie, W., Zhang, D., Poppe, R., Veltkamp, R. C., Li, B., & Yuan, J. (2019). A survey of variational and CNN-based optical flow techniques. *Signal Processing: Image Communication*, *72*, 9-24.

[20] Zhang, S., Wu, R., Xu, K., Wang, J., & Sun, W. (2019). R-CNN-based ship detection from high resolution remote sensing imagery. *Remote Sensing*, *11*(6), 631.

[21] Michele, A., Colin, V., & Santika, D. D. (2019). Mobilenet convolutional neural networks and support vector machines for palmprint recognition. *Procedia Computer Science*, *157*, 110-117.

[22] Gao, F., Wu, T., Li, J., Zheng, B., Ruan, L., Shang, D., & Patel, B. (2018). SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis. *Computerized Medical Imaging and Graphics*, *70*, 53-62.

[23] Salahat, E., & Qasaimeh, M. (2017, March). Recent advances in features extraction and description algorithms: A comprehensive survey. In *2017 IEEE international conference on industrial technology (ICIT)* (pp. 1059-1063). IEEE.

[24] Thanh, D. N., Erkan, U., Prasath, V. S., Kumar, V., & Hien, N. N. (2019, April). A skin lesion segmentation method for dermoscopic images based on adaptive thresholding with normalization of color models. In *2019 6th International Conference on Electrical and Electronics Engineering (ICEEE)* (pp. 116- 120). IEEE.

[25] Widodo, C. E., Adi, K., & Gernowo, R. (2020, April). Medical image processing using python and opencv. In *Journal of Physics: Conference Series* (Vol. 1524, No. 1, p. 012003). IOP Publishing.

[26] Chandan, G., Jain, A., & Jain, H. (2018, July). Real time object detection and tracking using Deep Learning and OpenCV. In *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 1305-1308). IEEE.

[27] Wang, W., Li, Y., Zou, T., Wang, X., You, J., & Luo, Y. (2020). A novel image classification approach via dense-MobileNet models. *Mobile Information Systems*, *2020*.

[28] Ginja, S., Gallagher, S., & Keenan, M. (2019). Water, sanitation and hygiene (WASH) behaviour change research: why an analysis of contingencies of reinforcement is needed. *International journal of environmental health research*, 1-14.

[29] Pool, E. R., Pauli, W. M., Kress, C. S., & O'Doherty, J. P. (2019). Behavioural evidence for parallel outcome-sensitive and outcome-insensitive Pavlovian learning systems in humans. *Nature human behaviour*, *3*(3), 284-296.

[30] Shullani, D., Fontani, M., Iuliani, M., Al Shaya, O., & Piva, A. (2017). VISION: a video and image dataset for source identification. *EURASIP Journal on Information Security*, *2017*(1), 1-16.