Journal of Information System and Technology Management JISTM
EISSN: 0128-1666
**Volume 8 Issue 32 (September 2023) PP. 36-52**
**DOI: 10.35631/JISTM.832003**

**JOURNAL OF INFORMATION SYSTEM AND TECHNOLOGY MANAGEMENT (JISTM)**
www.jistm.com

# GESTURE RECOGNITION FOR INTERACTIVE DASHBOARD: A 3D CONVOLUTIONAL NEURAL NETWORKS WITH LONG-SHORT-TERM-MEMORY APPROACH

Kah Kien Chye[1], Zi Yuan Ong[1], Jun Kit Chaw[2*], Mei Choo Ang[2], Marizuana Mat Daud[2], Tin Tin Ting[3]

[1] Department of Mathematical and Data Science, Faculty of Computing and Information Technology, Tunku Abdul Rahman University of Management and Technology, Malaysia
Email: chyekahkien@gmail.com, cozyworker123@gmail.com
[2] Institute of Visual Informatics, Universiti Kebangsaan Malaysia, Malaysia
Email: chawjk@ukm.edu.my, amc@ukm.edu.my, marizuana.daud@ukm.edu.my
[3] Faculty of Data Science and Information Technology, INTI International University, Malaysia
Email: tintin.ting@newinti.edu.my
[*] Corresponding Author

**Article Info:**

**Article history:**
Received date: 23.07.2023
Revised date: 15.08.2023
Accepted date: 05.09.2023
Published date: 15.09.2023

**To cite this document:**

Chye, K. K., Ong, Z. Y., Chaw, J. K., Ang, M. C., Daud, M. M., & Ting, T. T. (2023). Gesture Recognition For Interactive Dashboard: A 3D Convolutional Neural Networks With Long-Short-Term-Memory Approach. *Journal of Information System and Technology Management, 8* (32), 36-52.

**DOI:** 10.35631/JISTM.832003



**Abstract:**

Gesture recognition technology has become increasingly popular in various fields due to its potential for controlling consumer devices, robotics, and even translating sign language. However, there is still a lack of practical applications in daily life, especially for consumer products. Previous studies primarily focused on gesture recognition and did not delve into exploring the potential applications of gesture recognition in human interaction. This work aims to address this gap by deploying a gesture recognition model into practical applications and exploring its potential in human interaction. The proposed model combines the use of 3D convolutional neural networks (3DCNN) and long-short-term-memory networks (LSTM) for spatiotemporal feature learning. Six common interactive dashboard gestures were extracted from the 20BN-Jester dataset to train the model. The validation accuracy reached 80.47% after ablation studies. Then, two test cases were conducted to evaluate the effectiveness of the model in controlling Spotify and the Artificial Intelligence of Things (AIoT) smart classroom. The response time for each recognition was approximately 100 to 150 milliseconds. The proposed research aims to develop a simple yet efficient model for hand gesture recognition, which has the potential to be applied in various fields beyond just music and smart classrooms. With the growing popularity of gesture recognition technology, this study contributes to the advancement of practical applications and product innovations that can be integrated into daily life, thereby making it more accessible to the general public.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 8 Issue 32 (September 2023) PP. 36-52
DOI: 10.35631/JISTM.832003

**Keywords:**

Gesture Recognition; Deep Learning; Convolutional Neural Networks (CNN); Long Short-Term Memory (LSTM); Interactive Dashboard; Product Innovation

## Introduction

Human-computer interaction has expanded in various ways as a result of the rapid advancement of computer technology, from speech recognition to motion tracking and gesture detection (Iqbal, Naqvi, Khan, Khan, & Whangbo, 2022; Shera et al., 2021). Customer satisfaction towards products has undergone changes following the pandemic (Seap et al., 2022). In recent years, the market for gesture recognition has expanded quickly, particularly post-COVID-19 pandemic, where touchless interaction is much desired. Accurate real-time recognition and transmission of hand action signals are necessary for an interactive dashboard with hand tracking to improve the interactive experience. A developed and useful model is therefore necessary for widespread application and business goals. Gesture recognition is anticipated to displace current technology in the future. Future smartphones, tablets, desktops, laptops, and Internet-of-Things (IoT) gadgets will likely integrate more sophisticated capabilities that mix air gestures and radar sensors in favor of more conventional input methods like the mouse, keypad, and touch (Seiger, Kühn, Korzetz, & Aßmann, 2021). The idea of space-interval gesture manipulation, which is frequently shown in science fiction movies, is becoming more and more plausible. Therefore, our work aims to validate a gesture recognition model to be deployed into practical usage, namely interactive dashboards for Spotify and Artificial Intelligence of Things (AIoT) based smart classrooms.

### *Hand Gesture Recognition*

Existing hand gesture recognition techniques can be classified into two categories: vision-based and non-vision-based. The former usually uses cameras to capture color and depth, while the latter uses sensor-based methods to capture movements such as using data gloves. In this study, we focused on vision-based approaches only because Rautaray et al. (Rautaray & Agrawal, 2015) have examined earlier studies on the use of hand gestures in activities involving human-computer interaction and emphasized the significance of data processing in vision-based hand gesture identification. For the vision-based approach, hand gestures can be recognized as long as they can be seen via the cameras or sensors, no matter they are dynamic or static. A hand gesture recognition algorithm can be developed based on convolutional neural networks (CNN) and recurrent neural networks (RNN) (Núñez, Cabido, Pantrigo, Montemayor, & Vélez, 2018; Obaid, Babadi, & Yoosofan, 2020). Since gesture recognition includes image sequences, long short-term memory (LSTM) is also a popular approach for such problems (Y. Zhu, Xie, Zhang, & Li, 2023). The existing gesture recognition approaches may have added new features such as skeleton features (Núñez et al., 2018), contextual information (Wang, Gao, Song, & Shen, 2017) or others (Jiang & Chen, 2018; G. Zhu, Zhang, Shen, & Song, 2017) to improve the performance because they are not designed for human interaction systems. They may not capture all the subtle nuances and variations in gestures for the interactive dashboards.

Therefore, we extracted only specific gestures, including “Swiping Left”, “Swiping Right”, “Thumb Up”, “Thumb Down”, “Stop Sign”, and “No Gesture”, from the dataset. This selection ensured that the model was trained optimally for the interactive dashboard.. All the hand

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 8 Issue 32 (September 2023) PP. 36-52
DOI: 10.35631/JISTM.832003

gesture samples were made by numerous crowd workers and were performed over a variety of backgrounds. People's appearances vary greatly, while complex ambient occlusion and background scenes are also an added advantage of model training.

***Interactive Dashboard***

Gesture recognition technology's main target market at the moment is the consumer electronics industry, which also holds the highest market share. The rising demand from the automobile industry is responsible for this market's expansion (Shang, Liu, & Li, 2022). With more people using consumer electronics like computers, tablets, and smartphones, there is a bigger market for products that support gesture recognition technologies. Moreover, gesture recognition technologies have contributed to the enhancement of ergonomics in various consumer gadgets. Gesture recognition technology could become more widely available as a result of technological improvements. In addition, gesture recognition technology has potential uses in specialized industries like healthcare, where it can help with sanitation and disease prevention, or for people with disabilities who need alternative ways to connect with technology.

However, most companies and developers are having problems selecting the right tools, gesture recognition type and algorithms for their use-case. Therefore, the purpose of this study is to offer insights into the deployment of gesture recognition model on an interactive dashboard. It also aims to showcase how different types of gesture recognition perform in different situations, as certain types may be more suitable for specific use-cases. Additionally, we presented a viable use-case and showed how gesture recognition may be used practically in a real-world application. As a result, it may help to encourage the widespread use of gesture recognition technology, which may offer a fresh and effective method for consumer electronics.

The research is organized into 5 Sections, next Section 2 provides the literature review that focuses on gesture recognition models and technologies. Section 3 covers the gesture recognition model training and deployment of the interactive dashboard. In Section 4, the performance of our model and the overview of the two test cases of the dashboards are described, while in Section 5, the conclusion, limitation and future work are presented.

## Literature Review

Over the years, new studies have consistently emerged in the large subject of hand gesture detection. Nevertheless, realistic scenarios still present difficulties, and underlying intraclass variation and interclass similarities provide serious issues (Shang et al., 2022). Gesture acquisition, feature extraction, and classification algorithms are the common processes in hand gesture recognition.

Besides that, the different types of gesture recognition can be applied to various applications but yield different results. This makes it challenging to choose a single solution that applies to all situations. Furthermore, even within the same recognition system, the accuracy of various gestures can change. It may vary as well due to cultural and geographic differences. For dashboard interaction, while the set of movements and gestures is typically smaller, there are still challenges such as the need for specialized hardware, like stereo or time-of-flight cameras, which may not be readily available in consumer devices. Poor illumination and background noise might make it difficult to recognize motions in the real world, necessitating more

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 8 Issue 32 (September 2023) PP. 36-52
DOI: 10.35631/JISTM.832003

sophisticated motion detection. To maximize its potential for dashboard interaction systems, further research is needed to address these issues.

***Gesture Recognition Models***

The recognition process is affected by the proper selection of features parameters and a suitable classification algorithm. Various types of techniques can be used to implement the classification and recognition of images through machine learning. Prominent classifiers that used in gesture recognition include K-Nearest Neighbours (KNN), Support Vector Machines(SVM), Random Forest, Naïve Bayes, Logistic Regression and Multi-Layer Perceptron (Sharma, Mittal, Singh, & Awatramani, 2020).

Since gestures involve multiple frames to capture the essence of certain movements, models that are able to discriminate spatiotemporal features could outperform the others (Tsironi, Barros, Weber, & Wermter, 2017). Therefore, numerous neural network-based architectures have been proposed to learn spatiotemporal features for gesture recognition, which are inspired by deep learning breakthroughs in image recognition. Recent work using deep convolution neural networks (CNNs) has greatly improved the accuracy of dynamic hand gesture recognition and has proven useful for gesture recognition in the challenging lighting conditions (Jiang & Chen, 2018). CNN is a big step forward in image recognition. They are commonly used to analyze visual imagery and are often working behind the scenes in image classification. In the past, the extension to the video case for CNN remained an open issue. Baccouch et al. (Baccouche, Mamalet, Wolf, Garcia, & Baskurt, 2011) expanded CNN to the 3D convolution neural networks (3DCNN) to distinguish human actions using a depth video dataset, which can learn spatiotemporal features for long short-term memory networks (LSTM) automatically. Ohn-Bar et al. used the VIVA dataset (Ohn-Bar & Trivedi, 2014) to create an effective system for dynamic hand gesture recognition system using 3DCNN and LSTM. With the fine-tuned 3DCNN, excellent spatiotemporal features are learned automatically to characterize the hand movements video (Tran et al., 2020). Teja et al. (Reddy, Karri, Dubey, & Mukherjee, 2019) have also utilized 3DCNN for recognizing micro-expressions in videos. Figure 1 depicts the 3DCNN architecture proposed by Tran et al. (Tran et al., 2020). The network comprises six 3D convolutional layers (Conv1, Conv2, Conv3, Conv4, Conv5, and Conv6) in which Conv1 can receive video as input, three max-pooling layers (pool1, pool2, and pool3) to downscale the feature map, and two fully connected layers (fc1, fc2) before SoftMax output.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 8 Issue 32 (September 2023) PP. 36-52
DOI: 10.35631/JISTM.832003

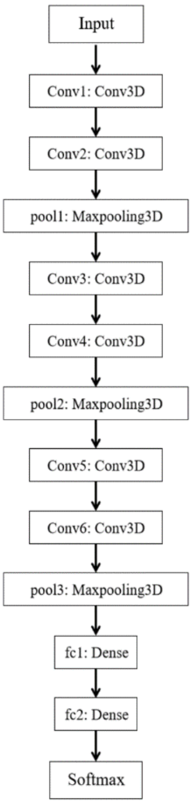


**Figure 1: Block Diagram of 3DCNN (Tran et al., 2020)**

Similarly, Molchanov et al. (Molchanov, Gupta, Kim, & Kautz, 2015) have achieved significant progress in the development of dynamic hand gesture recognition using 3DCNN. By conducting thorough evaluations, it has been demonstrated that combining low- and high-resolution sub-networks can significantly enhance classification accuracy. Moreover, the proposed data augmentation technique has been found to play a crucial role in achieving superior performance. The 3DCNN has a clear advantage over its 2D version because it performs convolutions in both the time and space dimensions to collect motion information. The model is also highly adaptive to huge datasets because it can directly take complete video frames as input without any prior preparation.

Gesture recognition, however, comes with its own sets of challenges to overcome, before the technology fully matures and is suitable for the wider market. These include problems in methodology, system building, interpretation of gestures, and the problem of applying it for practical applications. Other than that, challenges in recognizing gestures also arise due to environmental factors such as inadequate lighting, variations in users' skin color, and difficulties in detecting edges (Yasen & Jusoh, 2019).

***Deployment of Gesture Recognition Technology***

This section of the literature explores gesture recognition systems that have been implemented for user applications. Sarkar et al. (Sarkar, Gepperth, Handmann, & Kopinski, 2017) developed a real-time gesture recognition system that can operate on a mobile device using a 3D sensor optimized for sensor usage. The model used a combination of RNN and LSTM to achieve fast

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 8 Issue 32 (September 2023) PP. 36-52
DOI: 10.35631/JISTM.832003
speed recognition with great generalization capacity, even when it was executed on a tablet. This development creates the potential for seamless interaction on mobile devices, such as smartphones or tablets, through gesture recognition. An detection approach was proposed to resolve crucial real-time constraints by encoding generic lightweight hand gesture features through Fisher Vectors (Avola, Cinque, De Marsico, Fagioli, & Foresti, 2020). This feature representation was then analyzed through a LSTM to fully exploit video sequences.

## Methodology

In this study, we proposed a 3D convolutional layer for short-term spatiotemporal features and a Convolutional LSTM (3DCNN-ConvLSTM) layer for long-term spatiotemporal feature learning to handle the hand gesture recognition task. We then employed SoftMax to predict the hand gestures as well as applying Adam optimizer to handle sparse gradients on noisy problems. Subsequently, the model was deployed to realize an interactive dashboard. The architecture of the proposed 3DCNN can be seen in Figure 2.

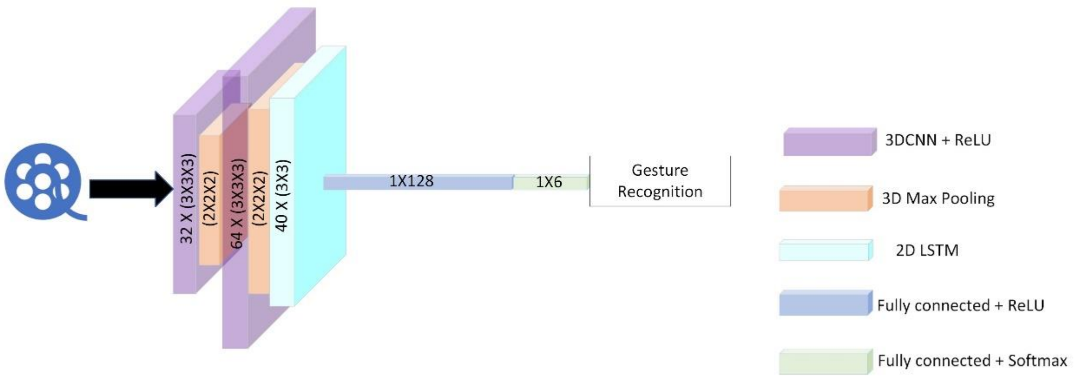


**Figure 2: Architecture of 3DCNN-ConvLSTM Model**

- This architecture is divided into the following layers.3DCNN layers, which enhance the recognition of three-dimensional and moving images. Each layer incorporates a filter that operates in three directions (x, y, z). Through 3D convolution, a convolutional map is generated, providing time and volumetric context for data analysis.
- MaxPooling3D layers are employed to downscale the 3D data, effectively reducing the dimensions of the video data.
- The dense layer consists of fully interconnected neurons that allow information to flow between them. The one positioned at the end of the network is responsible for converting the multidimensional matrix into a one-dimensional output vector. In our case, 6 output neurons for 6 gestures.

### *Model Training*

In Figure 3, the proposed system first read training and validation labels, allowing it to identify which video in the training and validation sample corresponds to which label.

The system then initialized the target labels, which were the hand gestures chosen for model training and prediction. The selected dataset for this study was the 20BN-Jester dataset (Materzynska, Berger, Bax, & Memisevic, 2019), which consisted of labeled video clips showing pre-defined hand gestures performed by humans in front of a laptop camera or webcam. The dataset was created by a large number of crowd workers and provides an opportunity to train robust machine learning models for recognizing human hand gestures. It

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 8 Issue 32 (September 2023) PP. 36-52
DOI: 10.35631/JISTM.832003

was selected because it is the largest video dataset available for gesture recognition and offers significant variability across actors performing the gestures. Six gestures were chosen from this dataset to be used in our study, which were "Swiping Left", "Swiping Right", "Thumb Up", "Thumb Down", "Stop Sign", and "No Gesture". These gestures were chosen because of their intuitiveness for interaction. Table 1 and Table 2 show the number of samples used for training and validation.

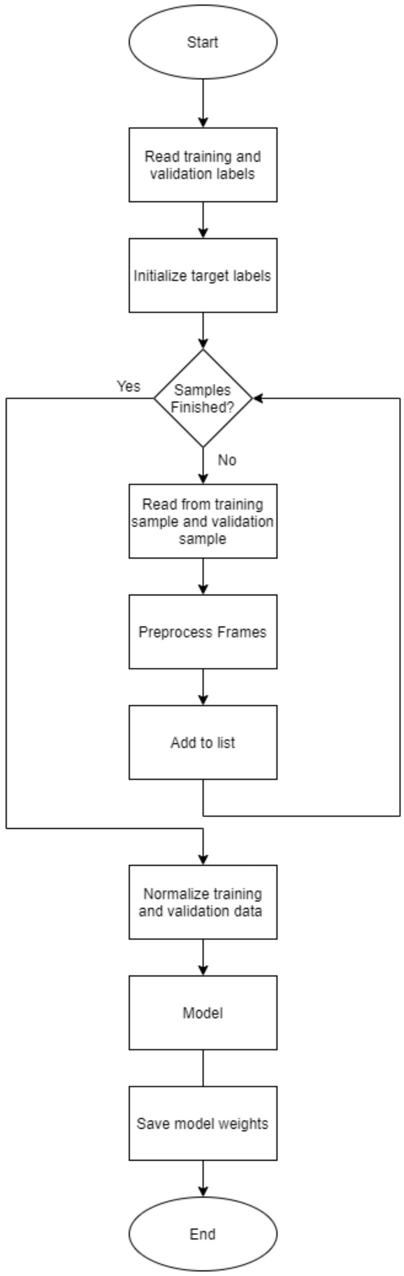


**Figure 3: General Flow Of The Overall System Design**

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 8 Issue 32 (September 2023) PP. 36-52
DOI: 10.35631/JISTM.832003

**Table 1: Overview Of The Training Samples**

| Gesture | Number of samples |
|---|---|
| Swiping Left | 1500 |
| Swiping Right | 1500 |
| Thumb Up | 1500 |
| Thumb Down | 1500 |
| Stop Sign | 1500 |
| No gesture | 1500 |

**Table: 2 Overview Of The Validation Sample.**

| Gesture | Number of samples |
|---|---|
| Swiping Left | 494 |
| Swiping Right | 486 |
| Thumb Up | 539 |
| Thumb Down | 536 |
| Stop Sign | 536 |
| No gesture | 533 |

The system would then load the training sample and validation sample directories and begin data pre-processing. We first unified the number of frames for each training before resizing the frame to 64 by 64 pixels and returned it as a grey image. To ensure that all videos had the same number of frames, a limit of 30 frames was used. Only the first 30 frames were extracted for use if the video is longer than necessary, and the last frame was duplicated if the video is shorter than required. A grayscale picture is one in which the only colors used are grayscale shades. After pre-processing, the frames were appended to two separate lists accordingly, one for the training sample and another one for the validation sample. Then, data was standardized by removing the mean and scaling to unit variance before data were entered into the model to improve model accuracy. The standardized training and validation samples acted as the input for model training. Lastly, the weights of the model were saved for use on the interactive dashboard.

***Interactive Dashboard***

The gesture recognition system was designed to interface for two test cases: 1) Spotify on a personal computer; 2) AIoT Smart Classroom. Therefore, it was designed to use gestures to control the various functions of Spotify and electrical devices. The model used in the dashboard was the same as the trained model training, a 3DCNN model combined with LSTM. The weights of the model were loaded from the model, therefore allowing the model to do prediction fast.

For Spotify, gestures were fed into the gesture recognition system to initiate actions such as play, stop, skip to next track, skip to previous track, volume up, and volume down, by taking advantage of Spotify’s keyboard shortcuts. The list of actions and their corresponding gestures are listed in Table 3 while the logical flow is shown in Figure 4.

Journal of Information System and Technology Management JISTM
EISSN: 0128-1666
Volume 8 Issue 32 (September 2023) PP. 36-52
DOI: 10.35631/JISTM.832003

**Table 3: Actions For Spotify Dashboard And Their Respective Gestures And Simulated Key Pressed**

| Action | Gesture | Simulated key pressed |
|---|---|---|
| Skip Backwards | Swiping Left | Right control + Left Key |
| Skip Forward | Swiping Right | Right control + Right Key |
| Volume Up | Thumb Up | Right control + Up Key |
| Volume Down | Thumb Down | Right control + Down Key |
| Stop/Play Tracks | Stop Sign | Left Mouse Button |
| Nothing | No gesture | Nothing |

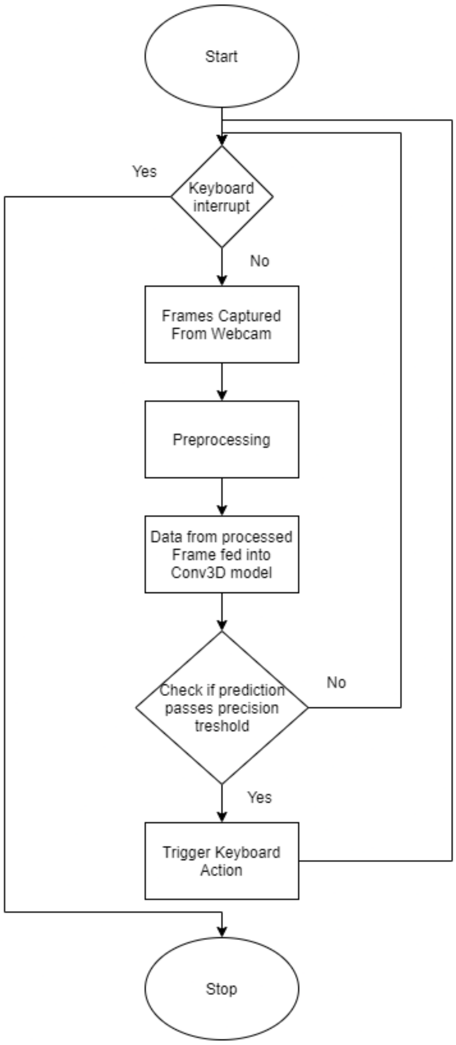


**Figure 4: General Flow of the Spotify Interactive Dashboard**

When the interactive dashboard was run, the model would return the list of precision for each possible gesture, and the gesture with the highest precision was chosen as the class of the gesture. After that, the precision of the most likely gesture was checked against a threshold of 0.95 precision for their respective gesture, and if it passed that threshold, it would trigger a keyboard action to interact with the Spotify dashboard. The system could be interrupted with

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 8 Issue 32 (September 2023) PP. 36-52
DOI: 10.35631/JISTM.832003

the key ‘q’.

For AIoT smart classroom, gestures were fed into the gesture recognition system to initiate actions such as open or lock door, on or off air conditioner, increase or decrease temperature and on or off light. In this AIoT Smart Classroom version, the sequence of actions is comparable to the previous version for the Spotify Dashboard, but with a key distinction - rather than using gestures to initiate a keyboard action, the gestures are used to prompt an insertion into the PostgreSQL for the air conditioning, door, or lighting, based on the specific gesture and the current status of the devices. The list of actions and their corresponding gestures are listed in Table 4 while the logical flow is shown in Figure 5.

**Table 4: Actions for AIoT Smart Classroom and Their Respective Gestures**

| **Action** | **Gesture** | **Data Sent** |
|---|---|---|
| Open/Lock Door | Swiping Left | True/False, opposite of current status |
| On/Off Air conditioner | Swiping Right | True/False, opposite of current status |
| Increase Temperature | Thumb Up | +1 to current temperature if still in range |
| Decrease Temperature | Thumb Down | -1 to current temperature if still in range |
| On/Off Light | Stop Sign | True/False, opposite of current status |
| Nothing | No gesture | Nothing |

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 8 Issue 32 (September 2023) PP. 36-52
DOI: 10.35631/JISTM.832003

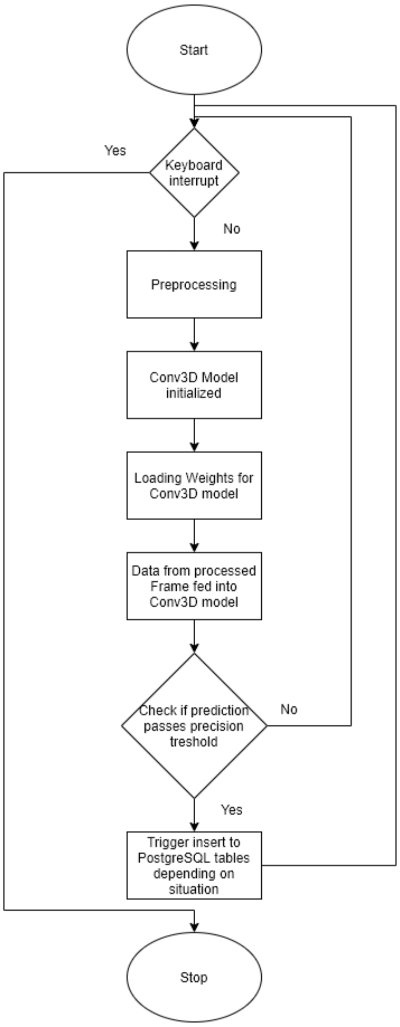


**Figure 5: General Flow of the Spotify Interactive Dashboard**

## Results and Discussions

The gesture recognition model was tested on the Jester dataset, a large-scale collection comprising approximately 5,000 instances of pre-defined hand gestures on average.On this challenging dataset, our model achieved validated accuracy of 80.47% which is shown in Figure 6. It leverages the effect of a hybrid model instead of using pure CNN and LSTM networks. The hybrid model takes advantage of CNN that can automatically extract discriminative features.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 8 Issue 32 (September 2023) PP. 36-52
DOI: 10.35631/JISTM.832003

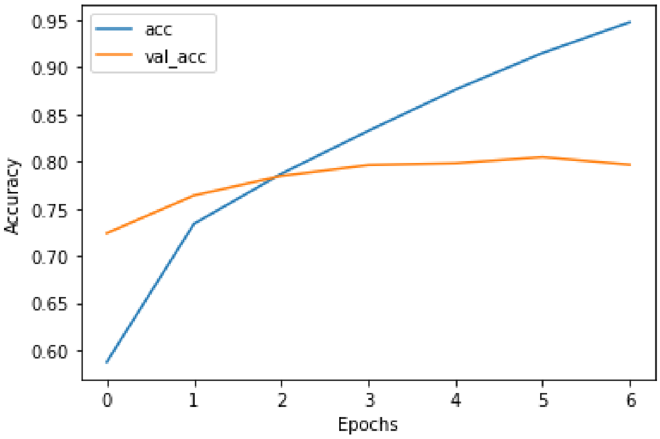


**Figure 6: Accuracy Curve (acc = training accuracy, val_acc = validation accuracy)**

Besides that, Figure 6 also shows that we obtained such results at the 6th training epoch. Results show that the model converged in 5th epoch and did not need more training. Thus, it would be the best point to stop at the 6th epoch. It is noticed that the training accuracy is increasing as the model learns. On the other hand, the validation accuracy seems to fall between 70% to 80%. This means that the model would overfit if it continues to train after the 6th epoch. To ensure the model is optimized, ablation studies were done which are shown in Table 5.

**Table 5: Results of Different Settings on the 6 Gestures from 20BN-Jester Dataset**

| Training setting | Validation Accuracy (%) |
|---|---|
| 3DCNN | 76.12% |
| 3DCNN + 2DLSTM | 78.58% |
| 3DCNN + 2DLSTM+ 3DMax Pooling | 80.47% |

Subsequently, this model was saved and used for the test cases. There is no particular UI for the Spotify interactive dashboard as it just detects the hand gestures according to the preset function in Table 3. However, the output of precision was shown in the terminal, which allows the user to grasp the general idea of the detection process. Each of the recognition only used from 100 to 150 milliseconds to show the prediction output on the terminal.

The test results for Spotify interactive dashboard are summarized in Table 6.

**Table 6: Test Results For Spotify Interactive Dashboard**

| Hand Gesture | Expected Result | Actual Result | Remarks |
|---|---|---|---|
| Swiping Left | Previous track | Previous Track | |
| Swiping Right | Next track | Next track | |
| Thumb Up | Volume increase | Volume increase | |

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 8 Issue 32 (September 2023) PP. 36-52
DOI: 10.35631/JISTM.832003

| | | | |
|---|---|---|---|
| Thumb Down | Volume decrease | Volume decrease | |
| Stop (when paused) | Play song | Play song | Since the 'Stop' gesture triggers the left mouse click, the mouse cursor must be positioned on the Play/Stop button to activate this action.. |
| Stop (when playing) | Stop song | Stop song | Since the 'Stop' gesture triggers the left mouse click, the mouse cursor must be positioned on the Play/Stop button to activate this action. |
| No gesture | Nothing | Nothing | "No gesture" is a class for any gesture that does not belong to the other five classes. In practice, it is activated only when no movement is detected by the camera. |

For AIoT smart classroom, the UI is shown in Figure 7.


**Figure 7: Overview of AIoT Smart Classroom Dashboard**

The UI was designed in a manner that is user-friendly and easy to understand. The top left section shows the status of the devices, with on/off indicators which are color-coded, with red for off and green for on. The top right section shows the current temperature of the air

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 8 Issue 32 (September 2023) PP. 36-52
DOI: 10.35631/JISTM.832003
conditioner, and finally, the bottom section shows the energy usage of the system in watt-per-hour over time. The refresh time of the dashboard can be customized to accommodate the desired frequency of displaying output changes. By adjusting the refresh time, the dashboard can update and reflect new information at specific intervals, providing a dynamic and real-time user experience. Similar to the interactive dashboard in Spotify, each gesture recognition process only took approximately 100 to 150 milliseconds to identify the user's gestures. Once the gesture was recognized, it was promptly sent to the corresponding control mechanism for further action shown in Table 7.

**Table 7: Test Results for AIoT Smart Classroom Interactive Dashboard**

| Hand Gesture | Expected Result | Actual Result | Remarks |
|---|---|---|---|
| Swiping Left (when door locked) | Open door | Open door | |
| Swiping Left (when door open) | Lock door | Lock door | |
| Swiping Right (when air conditioner off) | On air conditioner | On air conditioner | |
| Swiping Right (when air conditioner on) | Off air conditioner | Off air conditioner | |
| Thumb Up | Temperature increase | Temperature increase | The "Thumb Up" gesture will increase the temperature of the air conditioner until a maximum of 30°C, and the temperature can be changed regardless of the "on/off" status. |
| Thumb Down | Temperature decrease | Temperature decrease | The "Thumb Down" gesture will decrease the temperature of the air conditioner until a minimum of 16°C, and the temperature can be changed regardless of the "on/off" status. |

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 8 Issue 32 (September 2023) PP. 36-52
DOI: 10.35631/JISTM.832003

| | | | |
|---|---|---|---|
| Stop (when light off) | On light | On light | |
| Stop (when light off) | Off light | Off light | |
| No gesture | Nothing | Nothing | “No gesture” is basically a class for anything that doesn’t belong in the other five classes, and only triggers if there’s no movement detected on the camera. |

The caveat with these results is that sometimes the program would trigger the actions twice in a row, but for the most part erroneous detections did not occur due to the 95% confidence level threshold for triggering the action.

The biggest limitation in this study is the hardware requirements to achieve better and quicker training time if new actions were to be added into the interactive dashboard. With better hardware resources, the state-of-the-art of CNN and LSTM can be further utilized to train the model with higher recognition accuracy.

**Conclusion**

This study built a 3DCNN and LSTM hand gesture detection model which was trained with the 20BN-Jester dataset. For 6 different gestures, the final model showed satisfactory detection accuracy. The model was then integrated with the Spotify dashboard, enabling users to manage the dashboard with hand gestures. The interface system for an AIoT Smart Classroom System was also made by using the same gesture recognition model.

One of the limitations of this project is the limited number of trained gestures for the hand gesture recognition system. With only 6 trained gestures, the number of actions that can be mapped to these gestures and carried out is severely limited. Therefore, in future works, more gestures can be trained on the model to allow for more actions to be carried out and to further expand the functionality of this system.

Another limitation is that the current implementation of the dashboard control system relies on the dashboard being controlled to have an “inbuilt” keyboard shortcut that the control system triggers. Therefore, its function is limited to whatever the application allows the user to do with the keyboard. One possible future improvement is to further enhance its control scheme to adapt to whichever dashboard is currently being controlled, or to make it more flexible in its control scheme.

Overall, this study serves as a foundation for future improvements in gesture control systems. In the future, there is room for improvement in enhancing the accuracy of the gesture

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 8 Issue 32 (September 2023) PP. 36-52
DOI: 10.35631/JISTM.832003
recognition system and optimizing the implementation and design of the model to enable better performance of gesture predictions and make it more practical for real-world applications.

**Acknowledgment**
The authors would like to thank the Universiti Kebangsaan Malaysia for supporting this work through Geran Galakan Penyelidik Muda (GGPM-2022-063).

## References

Baccouche, M., Mamalet, F., Wolf, C., Garcia, C., & Baskurt, A. (2011). Sequential Deep Learning for Human Action Recognition BT - Human Behavior Understanding. In A. A. Salah & B. Lepri (Eds.) (pp. 29–39). Berlin, Heidelberg: Springer Berlin Heidelberg.

Iqbal, M. W., Naqvi, M. R., Khan, M. A., Khan, F., & Whangbo, T. (2022). Mobile Devices Interface Adaptivity Using Ontologies. *Computers, Materials \& Continua* . Retrieved from https://doi.org/10.32604/cmc.2022.023239

Jiang, S., & Chen, Y. (2018). Hand Gesture Recognition by Using 3DCNN and LSTM with Adam Optimizer BT. In B. Zeng, Q. Huang, A. El Saddik, H. Li, S. Jiang, & X. Fan (Eds.), *Advances in Multimedia Information Processing 2017* (pp. 743–753). Cham: Springer International Publishing.

Núñez, J. C., Cabido, R., Pantrigo, J. J., Montemayor, A. S., & Vélez, J. F. (2018). Convolutional Neural Networks and Long Short-Term Memory for skeleton-based human activity and hand gesture recognition. *Pattern Recognition*, 76, 80–94. Retrieved from https://doi.org/https://doi.org/10.1016/j.patcog.2017.10.033

Obaid, F., Babadi, A., & Yoosofan, A. (2020). Hand Gesture Recognition in Video Sequences Using Deep Convolutional and Recurrent Neural Networks. *Applied Computer Systems*, 25(1), 57–61. Retrieved from https://doi.org/doi:10.2478/acss-2020-0007

Ohn-Bar, E., & Trivedi, M. M. (2014). Hand Gesture Recognition in Real Time for Automotive Interfaces: A Multimodal Vision-Based Approach and Evaluations. *IEEE Transactions on Intelligent Transportation Systems*, 15(6), 2368–2377. Retrieved from https://doi.org/10.1109/TITS.2014.2337331

Rautaray, S. S., & Agrawal, A. (2015). Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, 43(1), 1–54. Retrieved from https://doi.org/10.1007/s10462-012-9356-9

Reddy, S. P. T., Karri, S. T., Dubey, S. R., & Mukherjee, S. (2019). Spontaneous Facial Micro-Expression Recognition using 3D Spatiotemporal Convolutional Neural Networks. In *2019 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). Retrieved from https://doi.org/10.1109/IJCNN.2019.8852419

Seap, V. M., Hoo, W. C., Shaznie, A., Ng, A. H. H., Chan, H. R., Pung, N. C., & Yew, L. K. (2022). Modeling Customer Satisfaction towards Private Hospitals during Covid-19 Pandemic in Klang Valley, Malaysia. *Res Militaris*, 12(5), 122–132.

Seiger, R., Kühn, R., Korzetz, M., & Aßmann, U. (2021). HoloFlows: modelling of processes for the Internet of Things in mixed reality. *Software and Systems Modeling*, 20(5), 1465–1489. Retrieved from https://doi.org/10.1007/s10270-020-00859-6

Shang, J., Liu, H., & Li, W. (2022). Human-Computer Interaction of Networked Vehicles Based on Big Data and Hybrid Intelligent Algorithm. *Wireless Communications and Mobile Computing*, 2022, 5281132. Retrieved from https://doi.org/10.1155/2022/5281132

Sharma, A., Mittal, A., Singh, S., & Awatramani, V. (2020). Hand Gesture Recognition using Image Processing and Feature Extraction Techniques. *Procedia Computer Science*,

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
**Volume 8 Issue 32 (September 2023) PP. 36-52**
**DOI: 10.35631/JISTM.832003**
173, 181–190. Retrieved from https://doi.org/https://doi.org/10.1016/j.procs.2020.06.022

Shera, A., Iqbal, M. W., Shahzad, S. K., Gul, M., Mian, N. A., Naqvi, M. R., & Khan, B. A. (2021). Blind and Visually Impaired User Interface to Solve Accessibility Problems. *Intelligent Automation \& Soft Computing*. Retrieved from https://doi.org/10.32604/iasc.2021.018009

Tran, D.-S., Ho, N.-H., Yang, H.-J., Baek, E.-T., Kim, S.-H., & Lee, G. (2020). Real-Time Hand Gesture Spotting and Recognition Using RGB-D Camera and 3D Convolutional Neural Network. *Applied Sciences*. Retrieved from https://doi.org/10.3390/app10020722

Tsironi, E., Barros, P., Weber, C., & Wermter, S. (2017). An analysis of Convolutional Long Short-Term Memory Recurrent Neural Networks for gesture recognition. *Neurocomputing*, 268, 76–86. Retrieved from https://doi.org/10.1016/J.NEUCOM.2016.12.088

Wang, X., Gao, L., Song, J., & Shen, H. (2017). Beyond Frame-level CNN: Saliency-Aware 3-D CNN With LSTM for Video Action Recognition. *IEEE Signal Processing Letters*, 24(4), 510–514. Retrieved from https://doi.org/10.1109/LSP.2016.2611485

Zhu, G., Zhang, L., Shen, P., & Song, J. (2017). Multimodal Gesture Recognition Using 3-D Convolution and Convolutional LSTM. *IEEE Access*, 5, 4517–4524. Retrieved from https://doi.org/10.1109/ACCESS.2017.2684186

Zhu, Y., Xie, M., Zhang, K., & Li, Z. (2023). A Dam Deformation Residual Correction Method for High Arch Dams Using Phase Space Reconstruction and an Optimized Long Short-Term Memory Network. *Mathematics*. Retrieved from https://doi.org/10.3390/math11092010