Journal of Theoretical and Applied Information Technology
31st May 2023. Vol.101. No 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

INTELLIGENT TOUCHLESS SYSTEM BASED ON GESTURE RECOGNITION

AISWARYA BABU[1], ZAHIRIDDIN RUSTAMOV[2] , SHERZOD TURAEV[3]

[1] Research Assistant, College of Information Technology, UAE University, UAE.

[2] Graduate. Faculty of Computer Science & Information Technology, University of Malaya, Malaysia

[3] Associate Professor, College of Information Technology, UAE University, U.A.E

E-mail: [1]aishdadu@gmail.com, [2]zakisher@gmail.com, [3]sherzod@uaeu.ac.ae

ABSTRACT

In our rapidly advancing technological era, every industry is experiencing its revolution. As we navigate the challenges the current pandemic presents, there is a heightened interest in solutions that facilitate social distancing and contactless interactions. To address this challenge, we propose the development of an interactive and innovative platform that allows users to navigate through hand gestures. This touchless system can be customized to meet various needs and utilizes a set of standard hand gestures for simplicity and ease of use and can be implemented in multiple sectors such as airports, banking, retail, restaurants, and so on. To demonstrate the system's potential, we have created a mobile food ordering application that uses hand gestures as the primary means of interaction and uses a set of standard hand gestures to promote simplicity, familiarity, and user accessibility. This study will develop a mobile food ordering system to illustrate the proposed gesture-based touchless system. To build our gesture recognition model, we collected a dataset of common hand gestures by scraping images from the web. We then trained our models using the Efficient Net-Lite [0-4] algorithms, leveraging transfer learning and pre-trained deep learning models to reduce computational demands. We utilized transfer learning and pre-trained deep learning models to reduce the time and computational resources required for training. The trained models were evaluated using the mean average precision (mAP) and inference time and then converted into a lightweight format, TensorFlow Lite, for use on mobile devices such as kiosks for the mentioned scenario. Our evaluation results revealed that all the trained models achieved an mAP of 82% or higher, with the most complex model, EfficientNet-Lite4, reaching 87%. However, the inference time for the trained models was significantly longer, ranging from one to ten seconds. To balance performance and inference time, we chose the EfficientNet-Lite0 model with an inference time of just half a second for our hand gesture-based touchless system. This model provides an adequate level of accuracy for our hand gesture-based touchless system while minimizing any lag or delay that could impact user experience. In summary, our proposed system is a cutting-edge, user-friendly solution that meets the need for contactless interactions and social distancing. Using standardized hand gestures, we have created a platform that is intuitive and accessible for users. Our system has the potential to offer significant benefits across a wide range of industries and applications in the modern era.

Keywords: ***Machine Learning, Artificial intelligence, Hand gestures, contactless, gesture-based***

1. INTRODUCTION

The fast-paced world we live in thrives on modernizing the techniques that will shape the upcoming future. The current technology is at the peak of its ever-evolving phase, and innovative solutions with a comprehensive approach are significantly accepted. This paper proposes the development of a smart interactive platform designed according to the social norms of maintaining social distancing in this pandemic period. The Hand Gesture-Based Smart Touchless System offers a baseline platform that utilizes a real-time gesture-based solution that can be customized to fulfill the varying requirements set forth by each targeted industry. The system is designed to use a set of standard hand gestures for simplicity, familiarity, and user accessibility. This allows the platform to comply with all the standard UI design standards, along with increasing the usability of this platform.

In the restaurant industry, the inclusion of technology increases the dining experience of its customers. This feature will fascinate the targeted crowds and give them a memorable dining

experience. The food ordering application thus fulfills the restaurant industry's requirements and stands out with its explicable experience.

This paper will focus on developing a mobile food ordering system that illustrates the proposed gesture-based touchless system. The motivation behind the idea of our application is three-fold: digitalization, minimal social contact, and great user experience. The food ordering application is designed to inculcate the gesture-based approach that provides the revolutionized extension of the current food ordering kiosks. The application is guided through a set of minimalistic design features and gestures that are used to navigate and order the desired item securely and reliably. We use out-of-the-box machine learning algorithms to create gesture recognition models that provide a cost-effective, innovative solution.

The gesture recognition model comprises the EfficientDet architecture, a subset of the MobileNet Architecture. This provides the baseline for our gesture recognition model employed on the mobile interface application. This paper also provides the efficiency of various models utilizing the EfficientDet architecture based on the chosen gestures.

2. RELATED WORKS

In the stake of recent pandemic events, there was a significant surge in the usage and implementation of technology in our daily routine.[1] The pandemic period resulted in a virtual lifestyle where social distancing became normal. This led to the usage of virtual-based solutions in each of our daily routines. All sectors were vastly affected due to this necessary lifestyle change and were improvised to meet their demands. One of the solutions to overcome the challenges faced was the implementation of contact-free technological solutions. [2]

Contact-free solutions encouraged the implementation of virtual working patterns, virtual shopping, virtual gaming, and so on. The on-site solutions included complying with the social distancing norms and limited entry to control the epidemic situation. Even in the post-pandemic phase, one has difficulty adjusting back to normal. [3] implements a gesture-based interface for gaming in public places in a secure manner by adhering to social norms and also by minimizing contact.

One of the many affected sectors was the restaurant industry, whose main objective was providing an on-site dining experience. While complying with the standard social distancing protocols, a contact-minimizing solution proved essential to maintain the integrity of this sector. Digitized solutions were implemented to comply with these standards, such as food delivery using virtual means [4]. To improve the dining-in experience, contactless solutions such as gesture or voice-based solutions can be implemented. Hence, nowadays, restaurants have been adapting kiosks and technological means for ordering to minimize the contact and risk of infections.[5]

Multiple solutions within the restaurant industry have been discussed in detail in the previous years. Among a few are [6] discusses a gesture-based model used to communicate with special individuals in a restaurant setting, [7] provides an outlook on the usage of basic gestures for selecting within a digitized menu. [8] provides an IoT-based solution for restaurant billing utilizing hand gestures for navigation. Hence, developing a smart gesture-based food ordering system implemented via restaurant kiosks will offer a cost-effective, innovative solution that complies with the social distancing protocol.

Multiple machine-learning algorithms can be utilized to implement a gesture-based interface that links machine-learning models for gesture detection and the mobile interface. [9] uses a CNN-LSTM-based integrated model for detecting signs and gestures. Even though the model proved effective, it was deemed incompatible with the latest mobile interface. Another method for gesture detection is via image processing through a transfer learning-based image recognition method called Mobilenet-RF[10], which is compatible with the mobile user interface when linked. [11] and [12] also provide high-accuracy models with variants of Mobilenet architecture. EfficientDet, used in the proposed model, is a subset of MobileNet architecture and provides an efficient solution for object detection [13]. The other models showed background interference in the hand detection and gesture identification process.

Therefore, our model is based on the EfficientDet algorithm for hand-gesture detection, as it is also compatible with the mobile interface. This method provides an efficient gesture recognition model with better accuracy prospectus while implemented for targeted sectors.

3. METHODOLOGY

The proposed system methodology for the smart touchless system navigated through hand gestures is shown in Figure 1 and is described in detail in the following sections.

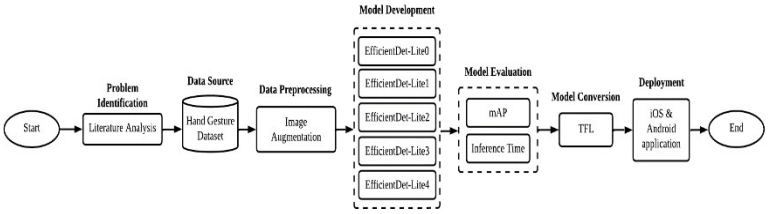

Figure 1: Research Methodology of This Study.

3.1 Dataset Description

This study utilizes two sources of images to construct the dataset for hand gesture modeling, with authors personally capturing and scrapping images from the web using Google Images. The constructed image dataset consists of eight classes for different hand gestures: option1, option2, option3, option4, back, ok, thumbs up, and fist. The description of each class is explained in Table 1.

Table 1: Description and Count of Each Class.

Class	Count	Description
option1	110	Index finger up
option2	110	Index and middle fingers up
option3	110	Index, middle, and ring fingers up
option4	110	Index, middle, ring, and little fingers up
back	110	All fingers up
ok	110	The ok sign or ring gesture; connecting the thumb and index into a circle while holding other fingers straight
thumbs up	110	Thumb up
fist	110	Clenched fist

The dataset contains 880 images, with each class comprising 110 images, as shown in Table 1. The resolution of the images differs greatly, with few images that exceed 2400 pixels × 2400 pixels. The sample images are illustrated in Figure 2.

This study performs data preprocessing to improve data quality before training the deep learning algorithms. As a result, this hand-picked study of images of different angles and resolutions ensures the algorithms are trained as thoroughly as possible.

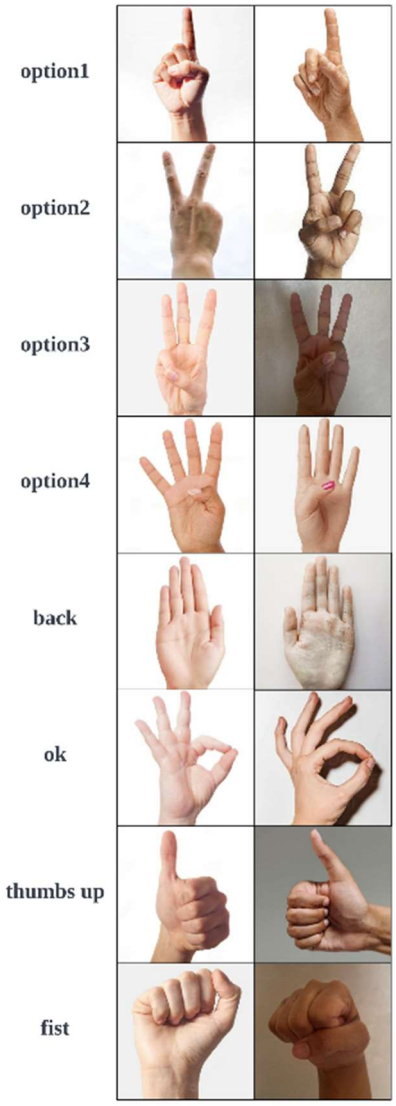

Figure 2: Sample Images for Each Class.

This study used an open-source tool, Labeling, to annotate the images in the constructed dataset, as shown in Figure 3. Consequently, we annotated 880 bounding boxes in the Pascal VOC format. The exact number of images within each class ensures that a class is not a majority or a minority, as it influences the model's performance. The dataset will be split into training and test sets, with 800 images utilized for training and 80 images for evaluation.

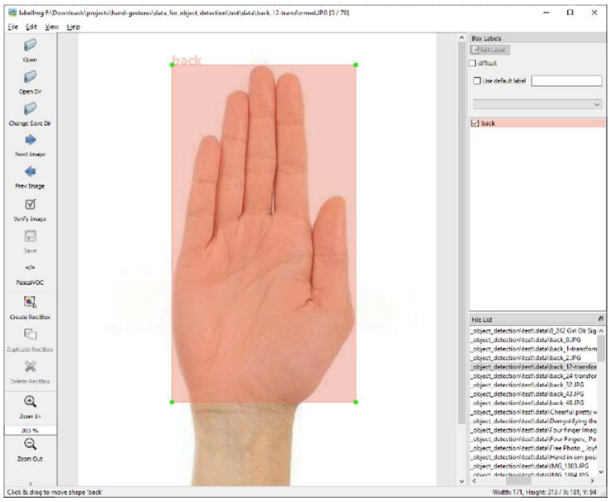

Figure 3: Process of Annotating Images using the Labeling tool.

This study will also perform image augmentation as a part of data preprocessing to train the models on a more robust dataset. We will use Albumentations, an open-source tool for performing image augmentation.[14] We will only perform augmentation on training images. Table 2 shows the applied transformations for image augmentation.

Table 2: Transformations Used for Image Augmentation

Transformation	Description	Parameters
SafeRotate	Rotates an image by 90°.	Limit: -90°, 90°
HorizontalFlip	Flips an image horizontally.	-
Random-Brightness-Contrast	Randomly changes the brightness and contrast of an image.	Brightness limit: 0.3, contrast limit: 0.3
PixelDropout	Drops some pixels of an image.	-

Following the image augmentation, each class's number of images and bounding boxes increased from 100 to 400, resulting in 3200 images in total. We will perform two instances of model development, with and without augmented images, to realize the effect of image augmentation.

3.2 Model Development

This study utilizes the TensorFlow Lite Model Maker library for model development as it automatically converts the trained models into a TensorFlow Lite format that can be run on mobile devices. Since this study aims to evaluate the proposed touchless hand gesture system on mobile devices, it is more appropriate for us to train lightweight algorithms that consider the computational limitations of mobile devices.

Therefore, this study will train the EfficientDet-Lite[0-4] algorithms on the constructed dataset. EfficientDet-Lite[0-4] is a family of mobile-friendly object detection models derived from the EfficientDet architecture. The main difference between the EfficientDet-Lite algorithms lies in their complexity, with EfficientDet-Lite4 being the most complex. As a result, the performance, model size, and latency differences between the algorithms.

We will train each algorithm for 50 iterations with a batch size of eight. This study utilized CUDA Toolkit 11.2 and CUDA Deep Neural Network (cuDNN) 8.1.0 to train the DL models using NVIDIA GeForce RTX 3080 10GB GDDR6X GPU.

3.3 Model Evaluation

This study utilizes the mean average precision (mAP) and inference time to evaluate the trained models. This study will use average precision (AP) interchangeably with mAP. The mAP is a commonly used evaluation metric for object detection tasks, with a higher mAP indicating a better-performing model. In particular, this study will adopt the detection evaluation metrics COCO uses. This study will consider the metrics AP@[.50: .05: .95] (i.e., AP), AP@.50 and AP@.75 for evaluation. The AP metric is stricter than the AP@.50 and AP@.75 as it considers intersection over union (IoU) over ten precision-recall pairs. The TensorFlow Lite Model Maker automatically evaluates the mAP of the models on the test data. The inference time indicates the time the model takes to infer the images, with lower inference time being favorable in real-time systems. This study will measure the inference time in milliseconds over 50 inferences and calculate by taking the average.

3.4 Mobile Application

This study will develop a cross-platform mobile application for the iOS and Android operating systems using the Flutter framework. The interface application complies with the UI design standards[15] and offers a user-friendly outlook accessible to all age groups.

4. RESULTS & DISCUSSION

This study aims to provide insight into developing a baseline smart touchless platform guided through hand gestures. This section will discuss the results of evaluating the food ordering application developed, teaching the characteristics of the smart touchless system navigated by hand gestures.

4.1 mAP Performance

This study utilizes the mAP and inference time metrics to evaluate the best-performing model for the touchless hand gesture-based system task. The models assessed are EfficientDet-Lite[0-4]. Table 3 shows the AP metrics for the trained models on non-augmented images for the test data. All the models achieved an mAP of 82% and above, with EfficientDet-Lite4 achieving the highest mAP of 87%. Although the EfficientDet-Lite0 reached the lowest mAP value, the difference between the top-performing model is insignificant, indicating that the least complex model performs almost as well as the most complex one. A notable observation is that there is no significant difference between the EfficientDet-Lite[2-4] models.

Table 3: The mAP Performance of Models Trained on Non-Augmented Images.

Model	AP	AP@.50	AP@.75
EfficientDet-Lite0	0.825	0.942	0.915
EfficientDet-Lite1	0.834	0.939	0.926
EfficientDet-Lite2	0.870	0.956	0.953
EfficientDet-Lite3	0.864	0.949	0.947
EfficientDet-Lite4	0.874	0.951	0.947

Table 4 shows the AP metrics for the trained models on augmented images. Similarly, we can observe an increasing trend in mAP as the models' complexity increases. The lesser complex models, such as EfficientDet-Lite0, perform poorly in accurately predicting the bounding boxes. Nonetheless, the models perform adequately when the IoU threshold is less strict, such as in the case of IoU at 50% or 75% (i.e., AP@.50 and AP@.75).

Table 4: The mAP Performance of Models Trained on Augmented Images.

Model	AP	AP@.50	AP@.75
EfficientDet-Lite0	0.608	0.885	0.756
EfficientDet-Lite1	0.644	0.894	0.808
EfficientDet-Lite2	0.643	0.900	0.828
EfficientDet-Lite3	0.723	0.945	0.894
EfficientDet-Lite4	0.671	0.915	0.812

Figure 4 shows a grouped bar plot of mAP comparison between models trained with and without augmented images. We can observe that the mAP of each model is significantly lower than those models trained without augmented images. The decline in the AP could be due to the model's inability to recognize the patterns in the augmented data, as there may not be enough examples. Moreover, the more complex models, such as EfficientDet-Lite4, may require more iterations to learn the patterns in data. Nevertheless, the AP@.50 metric still yields a value higher than 88% for each model, indicating that the models can generalize to the different transformations of each hand gesture.

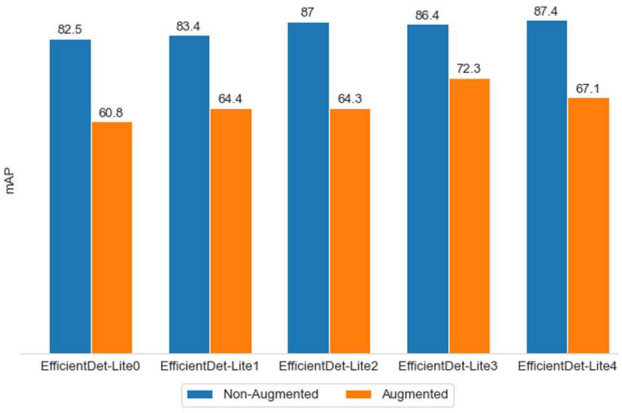

Figure 4: A Grouped Bar Plot of mAP Between the Models Trained with and Without Augmented Images.

4.2 Inference Time

This study calculated the inference time by taking the average time of 50 inferences by each model on a Xiaomi Redmi 9A mobile device, as shown in Table 5. We can observe that the inference time varies significantly between each model, with EfficientDet-Lite4 reaching up to 10 seconds to infer

a single image. The trend observed can be linked to the complexity of each model. As mentioned previously, the models differ in complexity; as such, more complex models require more computational resources from the mobile device for detection. The EfficientDet-Lite0 model achieves the fastest inference time on average, with less than a second.

Table 5: The Average Inference Time (in milliseconds) of Each Model.

Model	Average Inference Time (ms)
EfficientDet-Lite0	729.76
EfficientDet-Lite1	1315.78
EfficientDet-Lite3	2009.18
EfficientDet-Lite2	4085.24
EfficientDet-Lite4	10047.46

Figure 6 shows a scatter plot of the mAP performance of each model against the average inference time in seconds. We can observe a clear trade-off between the performance and latency, models with higher mAP resulting in higher inference durations. Although the EfficientDet-Lite2 takes lesser time to predict, it still outperformed a more complex model, EfficientDet-Lite3.

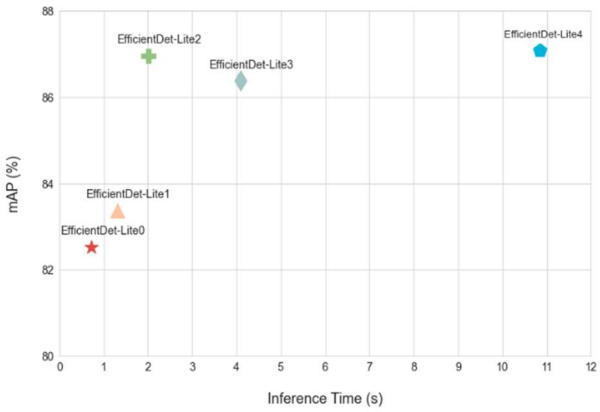

Figure 5: A Scatter Plot of Inference Time against mAP between the Models

4.3 Mobile Application

Figure 6. shows sample screenshots of the mobile application's user interface for evaluating the touchless hand gesture-based system.

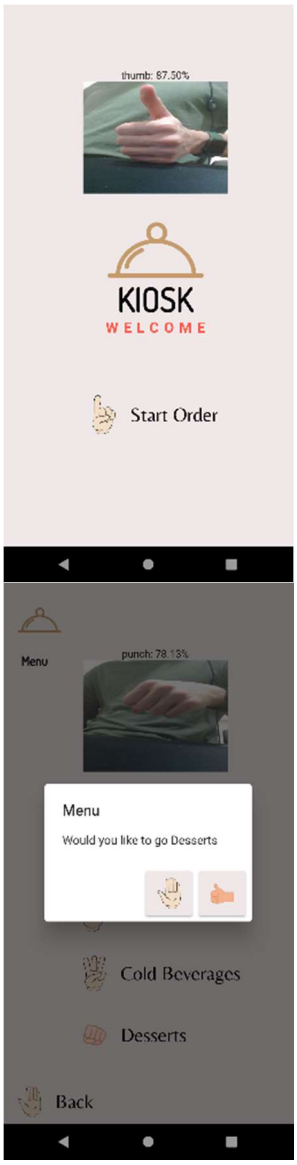

Figure 6: Screenshots of the Mobile Application for Evaluation.

5. CONCLUSION

The proposed hand-gesture-based smart touchless system provides an innovative outlook on how gestures can shape the customer service sectors while establishing social distancing. The food ordering smart touchless system provides an understanding of how the restaurant industry can utilize modern technologies and standard ordering mechanisms while providing an exemplary and memorable dining experience.

Future works should consider increasing the

number of images used for model development and evaluation while considering the environment in which the system is intended to be used. Moreover, the study should consider comparing the performance of the trained models against models based on hand key points detection.

ACKNOWLEDGMENT

The authors would like to thank the United Arab Emirates University for funding this work through UAEU-ZU Joint Research Grant G00003819 (Fund No.: 12R138), Emirates Center for Mobility Research.

REFERENCES

[1] G. George, K. Lakhani, P. P.-J. of Management, and undefined 2020, "What has changed? The impact of Covid pandemic on the technology and innovation management research agenda," *dash.harvard.edu*, vol. 57, no. 8, 2020, doi: 10.1111/joms.12634.

[2] H. Stevens and M. B. Haines, "Trace together: Pandemic response, democracy, and technology," *East Asian Sci. Technol. Soc.*, vol. 14, no. 3, pp. 523–532, 2020, doi: 10.1215/18752160-8698301.

[3] M. Roccetti, G. Marfia, A. S.-J. of V. C. and, and undefined 2012, "Playing into the wild: A gesture-based interface for gaming in public spaces," *Elsevier*, Accessed: Nov. 30, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1047320311001684

[4] L. Deksne, A. Kempelis, … T. S.-I., and undefined 2021, "Automated System for Restaurant Services.," *search.ebscohost.com*, vol. 24, pp. 15–25, doi: 10.7250/itms-2021-0003.

[5] S. Basar, "Developing a Concept for a Digital Restaurant System to Minimize Risk of Infection for Customers and Personnel," pp. 133–143, 2021, doi: 10.1007/978-3-030-66611-8_10.

[6] A. K.-I. J. of Multilingualism and undefined 2017, "Gesture-based customer interactions: deaf and hearing Mumbaikars' multimodal and metrolingual practices," *Taylor Fr.*, vol. 14, no. 3, pp. 283–302, Jul. 2017, doi: 10.1080/14790718.2017.1315811.

[7] I. Christian Susanto, K. Subaramaniam, and A. Samad bin Shibghatullah, "Gesturenomy: Touchless Restaurant Menu Using Hand Gesture Recognition," *alife-robotics.co.jp*, 2022, Accessed: Nov. 30, 2022. [Online]. Available: https://alife-robotics.co.jp/members2022/icarob/data/html/data/GS/GS4/GS4-3.pdf

[8] G. Kishore, … A. S.-2021 4th B., and undefined 2021, "Gesture Interfaced Restaurant Billing System using IOT," *ieeexplore.ieee.org*, Accessed: Nov. 30, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9487765/

[9] N. Basnin, L. Nahar, and M. S. Hossain, "An Integrated CNN-LSTM Model for Micro Hand Gesture Recognition," pp. 379–392, 2021, doi: 10.1007/978-3-030-68154-8_35.

[10] F. Wang, R. Hu, Y. J.-P. C. Science, and undefined 2021, "Research on gesture image recognition method based on transfer learning," *Elsevier*, Accessed: Nov. 30, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050921008243

[11] A. Alnuaim, M. Zakariah, … W. H.-C., and undefined 2022, "Human-Computer Interaction with Hand Gesture Recognition Using ResNet and MobileNet," *hindawi.com*, Accessed: Nov. 30, 2022. [Online]. Available: https://www.hindawi.com/journals/cin/2022/8777355/

[12] T. N. Abu-Jamie, P. Samy, and S. Abu-Naser, "Classification of Sign-Language Using MobileNet-Deep Learning," *Int. J. Acad. Inf. Syst. Res.*, vol. 6, pp. 29–40, 2022, Accessed: Nov. 30, 2022. [Online]. Available: https://philpapers.org/rec/ABUCOS-3

[13] A. Srikanth, A. Srinivasan, H. Indrajit, and N. Venkateswaran, "Contactless Object Identification Algorithm for the Visually Impaired using EfficientDet," *2021 Int. Conf. Wirel. Commun. Signal Process. Networking, WiSPNET 2021*, pp. 417–420, Mar. 2021, doi: 10.1109/WISPNET51692.2021.9419427.

[14] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and Flexible Image Augmentations," *Inf. 2020, Vol. 11, Page 125*, vol. 11, no. 2, p. 125, Feb. 2020, doi: 10.3390/INFO11020125.

[15] P. Reed, K. Holdaway, S. Isensee, … E. B.-I. with, and undefined 1999, "User interface guidelines and standards: progress, issues, and prospects," *academic.oup.com*, Accessed: Nov. 30, 2022. [Online]. Available: https://academic.oup.com/iwc/article-abstract/12/2/119/694538