Journal of Theoretical and Applied Information Technology

15th December 2023. Vol.101. No 23

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

DESIGN OF CONTEMPORARY MULTIVARIATE DATASET TO ASSESS THE QUALITY OF OBJECT, FACE AND PROXIMITY DETECTION IN ASSISTING THE VISUALLY IMPAIRED PEOPLE

S. SAJINI [1] , Dr. B. PUSHPA[2]

[1]Research Scholar, Department of Computer and Information Science, Faculty of Science. Annamalai university, Annamalai Nagar, Chidambaram, Tamil Nadu, India

[2]Assistant Professor / Programmer, Department of Computer and Information Science, Faculty of Science. Annamalai university, Annamalai Nagar, Chidambaram, Tamil Nadu, India.

E-mail: [1]sajinisheriff@gmail.com , [2]pushpasidhu@gmail.com

ABSTRACT

In recent years, advancements in Computer Vision have significantly impacted the development of assistive technologies for visually impaired individuals. The Major Objective of this research builds on the creation of a contemporary multivariate dataset designed to evaluate the quality of object, face, and proximity detection systems tailored to assist visually impaired individuals. The dataset incorporates diverse real-world scenarios, encompassing various environmental conditions and complexities commonly encountered by the visually impaired. It includes annotated images and accompanying ground truth data to facilitate the training and assessment of machine learning models for accurate object and face detection, as well as proximity estimation. The research work was based on the model designed using IoT enabled device and tested with 100 samples of visually impaired people. By leveraging this dataset, researchers and developers can enhance the performance of assistive technologies, ultimately improving the lives and independence of visually impaired individuals. The proposed dataset serves as a valuable resource for advancing the field of Computer Vision in the domain of accessibility and inclusive technology.

Keywords: *Object Detection; Face Detection; Proximity Detection; Multivariate Dataset; Visually Impaired; Computer Vision*

1. INTRODUCTION

Object detection technology uses smart computer programs and advanced technology to find and recognize different things instantly [1]. This included regular things like furniture, belongings, things that get the use of object finding and facial mortal faces and gathering important details about the people in a picture or video. Understanding the environment becomes more accessible for visually impaired individuals through the utilization of these technologies. It will make them more independent and confident in their daily activities. The use of object detection and facial detection technologies [2] can help blind people in many ways. It can detect things around them and help them know where it is. It can also help them mesh with others more easily. This thorough inquiry will delve into the elaborateness of technologies utilized for detecting objects and faces, and their specific purpose of aiding visually impaired individuals. Furthermore, the discussion will encompass the utilization, obstacles encountered, and prospective opportunities [3] that these technologies hold in enhancing the daily lives of visually diminished individuals.

Object detection technology can help blind people in their everyday lives. There are various applications for this technology, such as aiding in navigation and comprehending spatial orientation [4]. It can also assist with everyday activities. Some important needs of work include: Object discernment can help people who can't see well to transfer around without getting hurt. The technology can find things in the way, like furniture or curbs, and tell the user right away. This helps the

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

user not bump into things and go through places easily. Visually diminished people need to be able to find and know where objects are so can interact well with their surroundings. Object detection technology helps us find things like doors, chairs, tables, or our stuff. It makes it easier for us to find and use these things on our own. Adding Optical Character Recognition [5] to object detection systems, can help visually diminished pupils read printed text. The technology can recognize and understand written words on signs, labels, menus, or papers. It can then change this information into sound or touch output that is easy to access. Object observation can help blind persons safely move around in public transportation. The technology helps people use buses and trains more easily by telling them important information like bus or train numbers, signs at stations, or empty seats.

The utilization of computers in comprehending and interpreting visual information is the essence of object detection technology. It's a part of AI that helps machines [6] make sense of visual information. In pictures or videos, object identification involves identifying and pinpointing various entities. This technology is created by doing several steps like gathering information, teaching advanced computer systems, and using them to detect things happening in real time. The primary aspect to do when creating a model to detect objects [7] is to group a lot of labelled information. This information usually includes pictures that have labels showing where each thing is in the picture and what it is called. In a scenario where data is gathered to identify pedestrians, the images would exhibit designated boxes outlining the people, accompanied by labels identifying them. After gathering and labelling the information, complex computer systems are trained using advanced algorithms to understand and learn from it. Convolutional Neural Networks (CNNs) [8] are commonly utilized as they possess the ability to detect patterns in image data. In training, the neural network learns how to recognize special things about different objects and improves its settings to better detect them. After the model is trained and made better, it can be used on different devices like smartphones or tools specifically designed for people with vision problems. In real-time uses, the model analyzes videos or pictures [9] from the device's camera, finding and pinpointing objects in the picture. The objects that are found are then sent to the user through sound or touch, giving important information about what is around them.

Facial detection technology is a special kind of computer technology that looks for and distinguishes human faces in pictures or videos. This process uses advanced computer programs to recognize details of people's faces, emotions, and even specific people [10]. This technology is very important in helping visually impaired people have better social interactions and personalized experiences. Facial observation is all about finding a face in a picture or a video. Once a face is seen, the technology tries to find and mark important parts of the face like the eyes, nose, and mouth. These parts help understand how a person is feeling. Facial detection technology can also be used to find a specific person by comparing their face to a group of known people. This ability has a lot of potential in helping blind people recognize familiar faces. Facial detection technology helps visually impaired people [11] in many ways. It improves their social interactions and overall quality of life. There are some important uses. Facial detection technology can help people who can't see well by recognizing people and giving them useful information about them during social interactions. This means being able to know and remember people's faces, showing how they feel or look, and understanding how you know them. Understanding feelings by looking at facial expressions is a very important part of communicating. Facial detection technology can help people who can't see well to understand how others are feeling [12], so it can have more caring and knowledgeable conversations with them. Facial detection technology helps keep visually impaired people safe and secure. The technology helps identify people and evaluate possible dangers, which can make people more aware of their surroundings and feel safer.

The Major Objective of the Research Paper is to design a dataset that is multivariate in nature comprising three different variations including Object detection, Facial Detection and Proximity detection respectively. The Novelty of the work is that the hybridization of Object, Face and Proximity has not been achieved in any of the existing models. The Demographic details of the visually impaired persons are also considered for preparing the dataset. The scope of the research is confined to 100 samples collected from different visually impaired people to capture their images and store them in a dataset using image Capture techniques. The data samples were collected from clinical centers, and public organizations where visually impaired people were examined and evaluated with the performance of the Expert machine. Object observation and facial identification technologies can greatly help the visually diminished pupil to

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

understand and interact with their surroundings. Using these technologies can help the person with vision problems feel more included and able to get around on their own. The current progress in AI and computer vision shows that technology is improving to make the world a more inclusive place for everyone.

2. RELATED WORKS

Manuscripts Based on the problem being identified, the works of literature were studied in the context of the Object Detection as well as the Facial Detection process. These reviews were studied from recent works of Computer Vision in Object detection and Facial Recognition concepts respectively.

2.1 Review of Object Detection

In simple words, Wang, L., et al. (2023) [13] and others wrote this text. The text is about studying new methods for detecting fusion technology advancements in 2023. First, will give an overview of multi-modal 3D object exposure and explain the features of popular datasets and how are evaluated. Second, instead of using the usual way of classifying fusion methods into early, middle, and late fusion, examine and analyze all fusion techniques based on three aspects: how the features are represented, how are aligned, and how are combined. This helps us understand how these fusion methods are used fundamentally. Next, should thoroughly analyze the merits and demerits of both options and evaluate how well they perform in commonly used datasets. In simple terms, and outlines the main problems and areas of study needed to fully achieve multi-modal 3D object detection. Giakoumoglou, N., et al. (2023) [14] suggest a way to create fake datasets for object detection. It only needs a small dataset of the things want to detect and a bigger dataset of the background. The method called "Generate-Paste-Blend-Detect" uses Denoising Diffusion Probabilistic Models (DDPM) to create fake objects. These objects are then placed on a background image and blended with the surroundings to avoid bad image quality. Finally, an object observation model is used to identify the fake objects that were added. The new method was tested in farming to find whiteflies. It used the YOLOv8 model and had an accuracy of 0. 66 This way allows us to find specific things with less work and money.

Kaur, J., & Singh, W. (2022) [15] looks at object detection and talks about the different ways to detect objects. It reviewed current research on object observation and discussed seven questions about it, and have done different things for the research work. First, was analyzed different object detection techniques. Second, the preparation of a dataset and used a standard dataset that was already available. Third, they used the tools to mark important information in the dataset. Lastly, and evaluated how well our approach worked using certain measurement methods. Moreover, a comparison has been made to study and understand the unique characteristics of the proposed techniques. This comparison focused on their structures, methods of optimization, and approaches to training. Interior neural networks have greatly improved the presentation of object exposure. This research paper also talks about the future direction and ambition in object detection. Diwan, T., et al. (2023) [16] wrote a detailed review of single-stage object detectors, focusing on YOLOs, and discussed how these detectors work, their improvements in architecture, and their performance statistics. In addition, will explain the differences between two-stage and single-stage object detectors. It will also discuss various categories of YOLOs, applications using two-stage detectors, and future research ideas. Kang, J., et al. (2022) [17] give a detailed summary and comparison of the latest methods for using deep learning to detect objects in aerial photographs. In particular, our work can help us understand the latest improvements in detecting objects in pictures taken from above and the creation of sets of pictures taken from above that have not been fully learned before.

The creation of CODA, an intricate dataset, exposes a significant drawback in vision-based detectors. The dataset described by Li, K., et.al. (2022) [18] has 1500 real-world driving scenes that were carefully selected. Each scene has, on average, four unusual cases for objects, and includes more than 30 different types of objects. On CODA, the presentation of regular object detectors instructed on big self-driving datasets decreases to only 12. 8% in MAR Furthermore, and testing a very advanced object detector that is currently the best available, it discovered that it also struggles to accurately identify new objects in CODA. This indicates that creating a reliable system for autonomous driving that can fully understand its surroundings is likely still a long way off. It is the CODA dataset will make it easier for researchers to find ways to detect potential problems while autonomous cars are driving in the real world. Duan, R., et.al. (2022) [19] creates and shares a new collection of pictures for construction sites. It is called the Site Object Detection Dataset (SODA) and includes 15 different types of objects such as

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

workers, materials, machines, and layouts. Over 20,000 images were taking from various construction sites. These photos were taken in different situations, weather conditions, and stages of construction. It was taken from different angles and perspectives. The statistical analysis shows that the collection of data is extensive and diverse. Further analysis using two popular object observation procedure based on interior learning also shows that the dataset is workable, with the highest accuracy achieved at 81. 47% This research creates a big collection of pictures for the construction industry and provides a standard to measure how well different algorithms work. Kim, J. H., et.al. (2022) [20] fixed the explanatory of the SMD datafile and create a better version called SMD-Plus. This is used as a benchmark for DNN algorithms. It also suggested a technique to develop the SMD-Plus. To put it simply, the problem of having unequal numbers of different classes in the instructing data is fixed by copying and pasting images online. Additionally, the mix-up method is used along with the basic augmentation methods for YOLO-V5. The modified YOLO-V5 with the SMD-Plus works better at identifying and classifying things compared to the ordinary YOLO-V5, according to the test results. You can download the accurate information about the SMD-Plus and our test results.

The Dataset recommendation is to employ image-text models as a reliable method for identifying a broad range of objects. The approach of Minderer, M., et al. (2022) [21] involves utilizing a foundational Vision Transformer structure with minor alterations, wherein the model is trained to identify and comprehend images and text collectively while enhancing its detection capabilities through ongoing learning. The study on the scaling properties of this setup reveals that when the increase image-level pre-training and model size, and consistently see improvements in the downstream detection task. It improved the performance of detecting objects in text-dependent and image-dependent situations by providing suitable strategies and regulations. In the text, it talks about using a special kind of computer network to identify different types of objects. Ingle, P. Y., & Kim, Y. G. (2022) [22] help us determine if an object is normal or abnormal. The best technology to find either a handgun or a knife in a single camera view has an average accuracy of 84. 21% or 9020% After doing a lot of tests, the proposed method was able to detect various kinds of guns and knives with the highest accuracy of 97. 50% on the ImageNet data file and IMFDB. It accomplished an accuracy of 90. 50% on the open-image dataset, 93% on the Olmos dataset, and 90. 7% accuracy on the multi-view cameras. This device, which has limited resources, has performed well by detecting objects in a multi-view camera with an accuracy score of 85. 5%. Amudhan, A. N., & Sudheer, A. P. (2022) [23] a simpler computer program to find small items quickly. A new type of CNN was created to fulfil the needs mentioned earlier. The detection gets better when more information from the early layers and moved to the deeper layers. The network is farsighted because it works well with faraway things but is slow with close-up things. The new model's performance is evaluated by comparing it with the best models currently available. The evaluation is done on different public datasets like VEDAI, Visdrone, MS COCO, and OID. The new model can better find small objects, and there is a 32% increase in how quickly it works on Jetson Nano. To handle a complicated situation where signals are mixed up, a new strategy called visible-depth-thermal (VDT) SOD is suggested. This strategy uses three different types of images to help robots see and understand their surroundings. In the meantime, and created a machine to capture pictures in different lighting conditions. It also made a new set of examples for VDT SOD called VDT-2048 dataset. Several pictures will be used together to show the important parts. But it's impossible to avoid interference. In this study, Song, K., et al. (2022) [24] introduced a new method called hierarchical weighted suppress interference (HWSI) to combine information from different sources effectively. The detailed test results show that our method performs better than the most advanced methods

2.2 Review of Face Detection

Boyd, A., et al. (2023) [25] used human-created maps to teach the model to pay attention to important parts of images. The CAM mechanism was used to check how important the model thinks each part of the training batch is. It compared this importance with what humans think is important and punished the model when there were big differences. The results of our experiment on synthetic face detection showed that our approach, called CYBORG, greatly improved the accuracy. The tests of our approach on face images were created by six Generative Adversarial Networks using different classification network models. It also demonstrated that even when the training data by seven times or additional information like segmentation masks, the performance of CYBORG-trained models was still superior compared to other methods. Besides the main

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

outcome of this work, and noticed that when the clear indications of specific areas are added to the task of identifying fake faces, humans became better at classifying them accurately. This work has created a new topic of study on how to include human visual attention in practical loss functions. All the information, computer instructions, and previously trained models used in this research were provided along with this published document. IQA has yet to be utilized for the detection of images produced by AI. Even though real and fake images may look the same, most of the important information that can tell them apart is found in the frequency domain of the images. Using our understanding, it had taken out characteristics of image quality from both the frequency and spatial domains. The new method given by Kiruthika, S., & Masilamani, V. (2023) [26] has gotten the best result with an accuracy of 99% when different kinds of tests were done on standard sets of data and has also discussed how the proposed model can be applied to different situations and how can understand and explain the results it gives us. Naga, P., et al. (2023) [27] talk about the information used by researchers, what it consists of, and how it is created. Additionally, the main focus of this research is to collect the latest technologies and different approaches used on a set of data to achieve the most accurate results. However, there are still problems that need to be fixed. Sometimes, you can't use the same method to make a model with different emotions, datasets, training data, and testing data. Every model aims to get better and more accurate results than previous studies that use the confusion matrix. Wang, Z., et al. (2023) [28] suggest three different groups of face datasets with masks. These include a dataset to detect masked faces, a dataset to recognize masked faces in real-world scenarios, and a dataset to recognize masked faces that are created artificially. Additionally, it carried out comparison experiments on these three datasets for reference.

The planned model represented by Hangaragi, S., et al. (2023) [29] can find and understand the face by using Face mesh. Because of Face mesh, the model works in different situations like different lighting and surroundings. The model can also work with pictures of males and females of any age and race, even if they are not facing straight ahead. The attached wild face dataset and real-time images are used to teach the internal neural network in the model. When testing the image, if the features of the face in the image match with the features of any of the images used for training, the model will identify the person's name. If there is no match, the model will label the person as "unknown". The proposed model has accomplished an accuracy of 94. 23% in recognizing faces.

Jarraya, I., et al. (2023) [30] introduce a new type of model called Convolutional Neural Network for Animal Face Detection (CNNAFD). It also introduces a simplified version of CNNAFD called CNNAFD-MobileNetV2. Additionally, a new database called Tunisian Horse Detection Database (THDD) was created for detecting horse faces in Tunisia. CNNAFD used a method to change and refine images using different features and a new technique. A new type of convolutional layer called ANOFS-Conv is suggested which uses a method called ANOFS to choose the most important features. The ANOFS method helps to train the ANOFS-Conv layer effectively. CNNAFD uses assembled fully combined layers as a powerful classifier. The combination of CNNAFD and MobileNetV2 creates a new network called CNNAFD-MobileNetV2. This network improves the accuracy of classification and helps in making better detection choices. The new detector with the CNNAFD-MobileNetV2 network is successful and performs well compared to other detectors. It achieves high accuracy rates on different datasets. Liang, B., et al. (2023) [31] suggest using a special network called a deep map-guided triple network. This network has two main parts: a deep forecast network and a triplet feature extraction network. The map made by the deep forecast network can show the differences between real and not real faces in things like unevenness, differences in lighting, and blurriness. This helps in detecting deepfake images. No matter how much someone's face may look different in a deepfake, and believe real and not-real faces should have different hidden characteristics. Specifically, the two real faces (original and target) stay close together in the hidden characteristic space. However, the two pairs of real-fake faces (original-fake and target-fake) stay far apart from each other. Using this approach, it proposed a network that helps extract distinct deep features by reducing the distance between the original and target pairs, while increasing the distance between the original and fake (or target and fake) pairs. Our method has been proven to be better than other techniques by testing it on the Face Forensics++ and Celeb-DF datasets. Vijaya Kumar, D. T. T., & Mahammad Shafi, R. (2023) [32] suggests a new method for face detection called Region-based Fully CNN (R-FCN). The R-FCN is a type of shape that uses a special layer to calculate a score for predicting each region. This

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

helps to make the network faster and share the processing of certain parts of an image, known as Region of Interest (RoIs), which prevents the loss of important information in RoI-pooling. In this study, a new algorithm called GE-GWO is developed to improve face detection by optimizing the R-FCN structure. It used the WIDER face dataset along with the Face Detection Dataset and Benchmark (FDDB) to test and assess different methods. and recall) compared to existing methods.

Sheikh, B. U. H., & Zafar, A. (2023) [33] suggest a fast and automated system called RRFMDS that can detect if someone is not wearing a face mask in a live video. In the new system, a special tool is used to find faces quickly, and a customized tool is used to determine if a face has a mask on it. The system is light and doesn't need a lot of resources. It can work together with CCTV cameras that are already installed to find people who are not wearing face masks. The main reason for making this dataset was to create a system that can detect different kinds of face masks, no matter how it is positioned by them. The system can recognize three types of people (wearing masks incorrectly, wearing masks correctly, and not wearing masks) with an average accuracy of 99. 15% and 9781% on the training and testing data. On average, it takes about 0. 14201142 seconds for the system to do everything it needs to do with a single image, such as finding faces and analyzing them. Firdaus, T. M., et al. (2023) [34] are trying to gather correct examples of Indonesian faces for face detection methods. Identifying the facial features of Indonesian people is difficult because Indonesia has different tribes. On the other words, tribes in Indonesia can be divided into Malay tribes and tribes that are not Malay. So, this study will use a collection of Indonesian people's faces that have gathered ourselves. Then test them using a special algorithm called FaceNet. The study found that simple devices can still give accurate results. This study shows that researchers can get more accurate results by using their method, even when using their data.

Luo, S., et al. (2023) [35] analyzed the various elements that can affect the maximum IoU value for each face. Next, a simulation is done to match the anchor to determine the orbit of the face aspect ratio that should be sampled. Finally, the suggestion of a WARM method is to find more accurate positive anchors from real faces with different shapes. Also, it has a new module called the Receptive Field Diversity (RFD) module, which helps create different receptive fields for different aspect ratios. Then tested our method extensively on well-known benchmarks to prove that it works well in detecting faces with unusual shapes and sizes. Our method performs well on the WIDER FACE validation dataset with high average precision scores (easy: 0. 965, medium: 0955, hard: 0904) and also shows impressive generalization ability on the FDDB dataset. In the work from Liu, W., Hasan, I., & Liao, S. (2023) [36], the process of detecting pedestrians and faces is made easier by predicting their centre and size using convolutions. The new method does not require an anchor, making it easier to train and optimize. Although it is not complex in structure, this technology performs well in accurately identifying objects like pedestrians and faces in difficult tests compared to other systems.

3. MATERIALS AND METHODS

The Methodology to design a Multivariate Dataset in making a dataset with pictures to help people who can't see well in detecting objects and faces involves a few important steps. In the following, various steps were explained in making a set of data while highlighting the importance of training the model correctly for creating a Multivariate Dataset as shown in Fig.1.

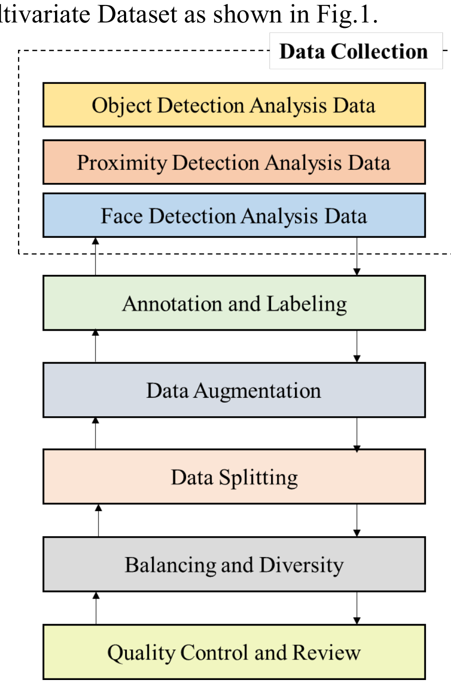


Fig.1. Different Stages of the Multivariate Dataset Collection Process for assisting Visually Impaired

The different stages as portrayed in Fig.1. for designing a Multivariate dataset for Visually Impaired people have been presented in different heads. 1. Data Collection: Collect many different samples that show objects and their faces that

15th December 2023. Vol.101. No 23

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

Visually Impaired people often go through. Make sure there are different lighting conditions, backgrounds, and a variety of objects and faces present in a detection environment. The goal is to make a dataset that looks like things that happen in real life. 2. Annotation and Labeling: Mark the important parts of the pictures, like objects and faces, using boxes. Make sure to put clear names or descriptions that show what kind of thing it is or who the person is if it's a face. Accurate and detailed annotations are very important for effectively teaching models. 3. Data Augmentation: Use different methods to increase the variety and durability of the dataset. Methods like turning, changing size, reversing, and changing colours can make more versions of the original pictures. This helps the model understand different situations better. 4. Data Splitting: divide the data file into smaller parts called training, validation, and testing sets. The training set is used to teach the model, the validation set helps make the model work better, and the testing set checks how well the model performs. 5. Balancing and Diversity: Make sure that there are equal amounts of different objects, faces, and distances in the dataset based on thresholds and limits. A balanced dataset helps the model learn better without favoring certain categories. This makes sure that the model can accurately identify various objects and faces. 6. Quality Control and Review: Check annotations carefully to make sure they are accurate and consistent. To keep the dataset accurate, fix any mistakes or differences quickly. These corrections are important for creating trustworthy models. The dataset creation is possible with the utilization of a Framework model designed and implemented using the sensor-enabled kit that brings all three major technologies of Computer Vision including Face detection, Proximity detection and Object detection. The Multivariate dataset was created using an Analytics sheet following the guidelines. In the last few years, technologies like object detection, facial detection, and proximity detection have become important solutions to help blind people have better lives. These advanced technologies, which use AI and computer vision, can help immediately by giving important information about things around us. This includes identifying objects, recognizing faces, and seeing obstacles nearby. By using sound or touch cues, these technologies help people who can't see to move around on their own and stay safe. To encompass the various parts of designing an analysis system, an analysis report was designed with demographic details and the performance of the machine in three variations including Object Detection, Proximity Detection and Face Detection as shown in Table 1

Table 1: Analysis Report Sample for Data Collection of Multivariate Dataset.

DEMOGRAPHIC ANALYSIS

NAME		**PLACE**	
AGE	17-24 24-30 31-50	**GENDER**	MALE FEMALE
COMPLEXION	FAIR BROWN DARK	**RACE**	BLACK WHITE AVERAGE
PPE – SPECS	1. YES 2. NO	**PPE-MOUTH MASK**	YES NO
PPE – HAIR MASK	YES NO	**PPE- EAR MASK**	YES NO

VISUAL ASSISTANCE EXPERT ENGINE ANALYSIS

FACIAL DETECTION	**TEST #1**	**TEST #2**	**TEST #3**	**TEST #4**	**TEST #5**	**TEST #6**	**TEST #7**	**TEST #8**	**TEST #9**	**TEST #10**
	0 – UNKNOWN 1 – CORRECT NAME 2 – WRONG NAME									

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

<table>
<tr><td rowspan="3">PROXIMITY DETECTION</td><td colspan="10">THRESHOLD – 12 MTS</td></tr>
<tr><td>TEST #1</td><td>TEST #2</td><td>TEST #3</td><td>TEST #4</td><td>TEST #5</td><td>TEST #6</td><td>TEST #7</td><td>TEST #8</td><td>TEST #9</td><td>TEST #10</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td colspan="10">0 – EXACT
1 – PROXIMAL
2 - WRONG</td></tr>
<tr><td rowspan="2">OBJECT DETECTION</td><td>OBJ #1</td><td>OBJ #2</td><td>OBJ #3</td><td>OBJ #4</td><td>OBJ #5</td><td>OBJ #6</td><td>OBJ #7</td><td>OBJ #8</td><td>OBJ #9</td><td>OBJ #10</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td colspan="10">0 – UNKNOWN
1 – CORRECT OBJECT
2 – WRONG OBJECT</td></tr>
</table>

As shown in Table 1., the major part of the dataset has been segregated into two major components including

1. Demographic Analysis
2. Visual Assistance Expert Engine Analysis

In Demographic Analysis, various demographic factors of the visually impaired like the following details in Table 2. have been collected.

Table 2: Demographic Details for Analyzing Visually Impaired People Center

NAME		**PLACE**	
AGE	17-24 24-30 31-50	**GENDER**	MALE FEMALE
COMPLEXION	FAIR BROWN DARK	**RACE**	BLACK WHITE AVERAGE
PPE – SPECS	YES NO	**PPE- MOUTH MASK**	YES NO
PPE – HAIR MASK	YES NO	**PPE- EAR MASK**	YES NO

Among the features in demographic details represented in Table 2., "Name" and "Place" were nominal features that couldn't be used in prediction as name and place have no serious impact on their predictions. The Features Age, Gender, Complexion, Race, Specs, Mouth Mask, Hair Mask, and Ear Mask as Personal Protective Equipment (PPE) were predictive features that accept a range of values as shown in Table 2. The Age has been collected in three levels including 1) 17-24 to indicate adolescent age groups, 2) 24-30 to indicate young people ready for work, and 3) 31-40 indicating middle-aged working people respectively. The Gender has been collected in two levels including 1) Male and 2) Female, Complexion in three levels including 1) Fair, 2) Brown and 3) Dark, and Race with three levels including 1) Black, 2) White and 3) Average to test whether the complexion and Race has an influence on visually impaired people respectively. The remaining PPE-based demographic variables are based on YES/NO type to indicate the presence or absence of the object in the face of the visually impaired. The first phase of dataset collection was initiated with Facial detection and included 10 tests on the same person as shown in Table 3.

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

Table 3: Facial Recognition Tests for Analyzing Visually Impaired People

FACIAL DETECTION	TEST #1	TEST #2	TEST #3	TEST #4	TEST #5	TEST #6	TEST #7	TEST #8	TEST #9	TEST #10
	1 – UNKNOWN 2 – CORRECT NAME 3 – WRONG NAME									

As shown in Table 3., there were columns to test every individual face of the Visually impaired Person in 10 consecutive tests and determine whether the machine was capable of recognizing the face with the correct name in all subsequent tests. The answers are recorded with responses of 1 (Unknown), 2 (Correct Name) and 3 (Wrong Name). The wrong name represents the wrong name value identified by the machine whereas the "Unknown" represents the machine couldn't recognize the face of the person.

The Proximity of the Face or object has been detected in the second level with a threshold distance of 12 meters from the position of the visually impaired. It also included 10 tests on the same person as shown in Table.4.

Table 4: Proximity Detection Tests for Analyzing Visually Impaired People

PROXIMITY DETECTION	THRESHOLD – 12 MTS									
	TEST #1	TEST #2	TEST #3	TEST #4	TEST #5	TEST #6	TEST #7	TEST #8	TEST #9	TEST #10
	1 – EXACT 2 – PROXIMAL 3 - WRONG									

As shown in Table 4., there were columns to test every object or face with their proximality with the Visually impaired Person in 10 consecutive tests and determine whether the machine was capable of recognizing the proximity in all subsequent tests. The answers are recorded with responses of 1 (Exact), 2 (Proximal) and 3 (Wrong). The wrong name represents the wrong object proximity identified by the machine whereas the "Proximal" represents the closeness of the machine with the object or person. The Exact is the right prediction within the exact threshold of 12 meters indicating the correct test of the object for the assistance of the visually impaired.

The Final phase of dataset collection initiated with the Object detection included 10 tests on the same object as shown in Table 5.

Table 5: Object Detection Tests for Analyzing Visually Impaired People

OBJECT DETECTION	OBJ #1	OBJ #2	OBJ #3	OBJ #4	OBJ #5	OBJ #6	OBJ #7	OBJ #8	OBJ #9	OBJ #10
	1 – UNKNOWN 2 – CORRECT OBJECT 3 – WRONG OBJECT									

As shown in Table 5., there were columns to test every individual object of the Visually impaired Person in 10 consecutive tests and determine whether the machine was capable of recognizing the object with the correct name in all subsequent tests. The answers are recorded with responses of 1 (Unknown), 2 (Correct Object), and 3 (Wrong Object). The wrong name represents the wrong name value identified by the machine whereas the "Unknown" represents the machine couldn't recognize the object.

4. PREDICTIVE ANALYTICS OF THE MULTIVARIATE DATASET

Object Detection technology is a tool that helps us find and figure out where things are in our environment. It can find and locate different things like furniture and personal items. This skill helps people understand their surroundings and make choices about where to go. On the other hand, facial detection technology is all about figuring out who people are by looking at their faces. It helps in social situations by giving us information about the people around us. Combining proximity detection helps us better understand our surroundings by identifying objects or people nearby. This adds to our overall understanding of the environment. The dataset was collected from 100 samples where the demographic distributions are given in different categories.

The Ages of the respondents were collected based on the samples from public locations with the following distributions given in Fig.2

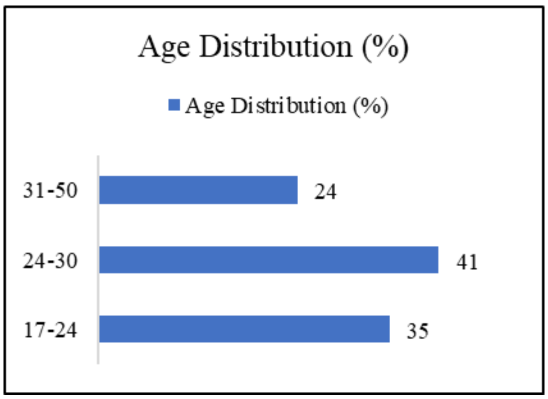


Fig.2. Age-based Distribution of the Multivariate Dataset

As shown in Fig.2., the age group of respondents of 17 to 24 were 35% 24 to 30 were 41% and 31 to 50 were 24% respectively. The Gender of the respondents was collected based on the samples from public locations with the following distributions given in Fig.3.

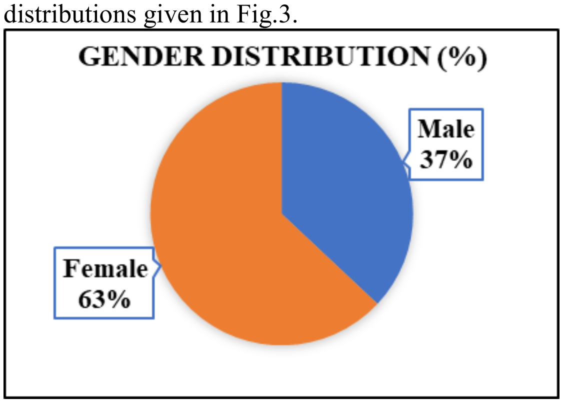


Fig.3. Gender-based Distribution of the Multivariate Dataset

As shown in Fig.3., the gender group of respondents of Male were 37% and female were 63% respectively. The Complexion of the respondents were collected based on the samples from public locations with the following distributions given in Fig.4.

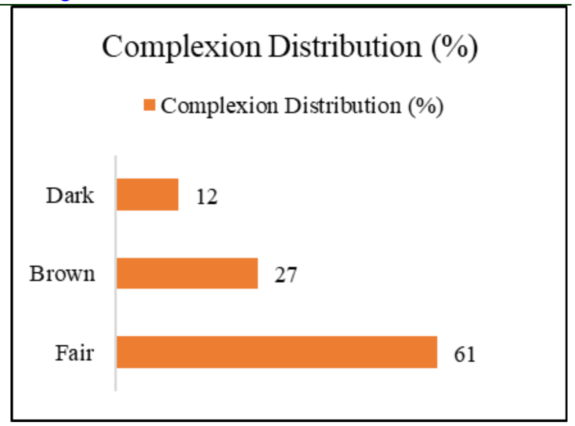


Fig.4. Complexion based Distribution of the Multivariate Dataset

As shown in Fig.4., the complexion group of respondents of Fair were 61% Brown 27% and Dark 12% respectively. The Race of the respondents were collected based on the samples from public locations with the following distributions given in Fig.5.

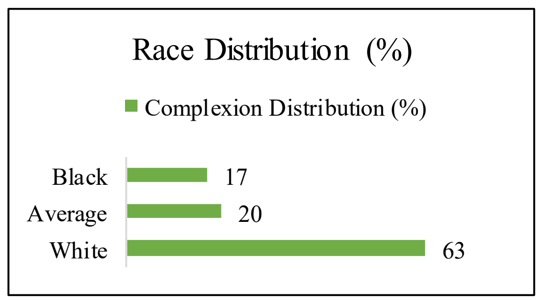


Fig.5. Race-based Distribution of the Multivariate Dataset

As shown in Fig.5., the Race group of respondents of white were 63% and Average were 20% and Black were 17% respectively. The remaining Personal Protective Equipment (PPE) based object components including Specs, Mouth, Hair Mask, and Ear of the respondents were collected based on the samples from public locations with the following distributions given in Fig.6.

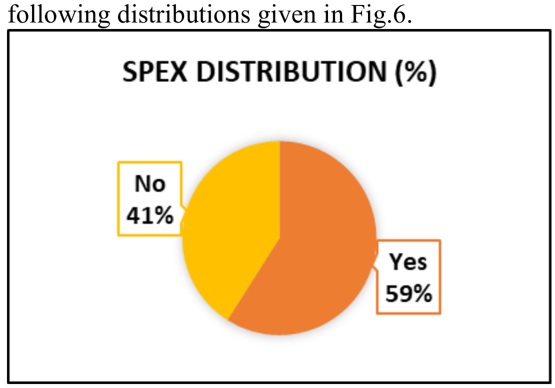


Fig.6(a). PPE-Specs-based Distribution of the Multivariate Dataset

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

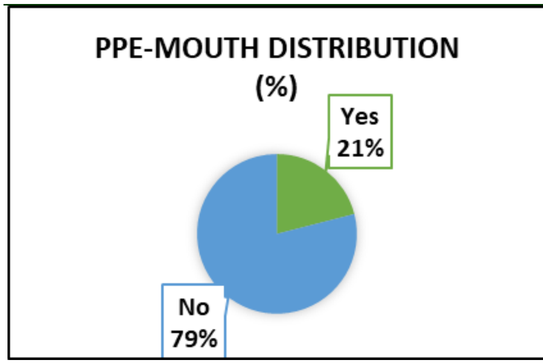


Fig.6(b). PPE-Mouth-based Distribution of the Multivariate Dataset

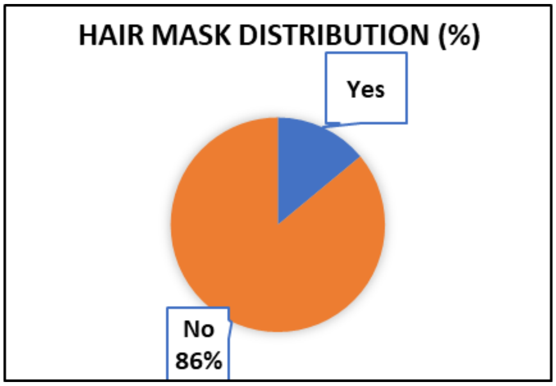


Fig.6(c). PPE-Hair Mask-based Distribution of the Multivariate Dataset

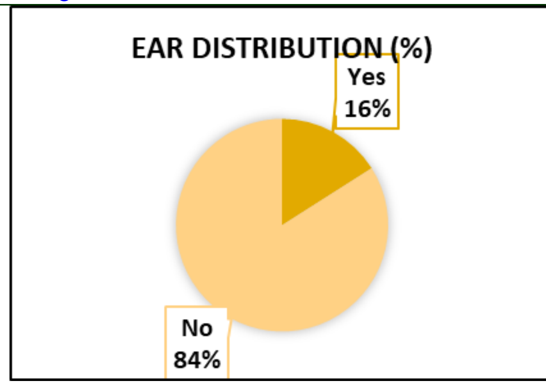
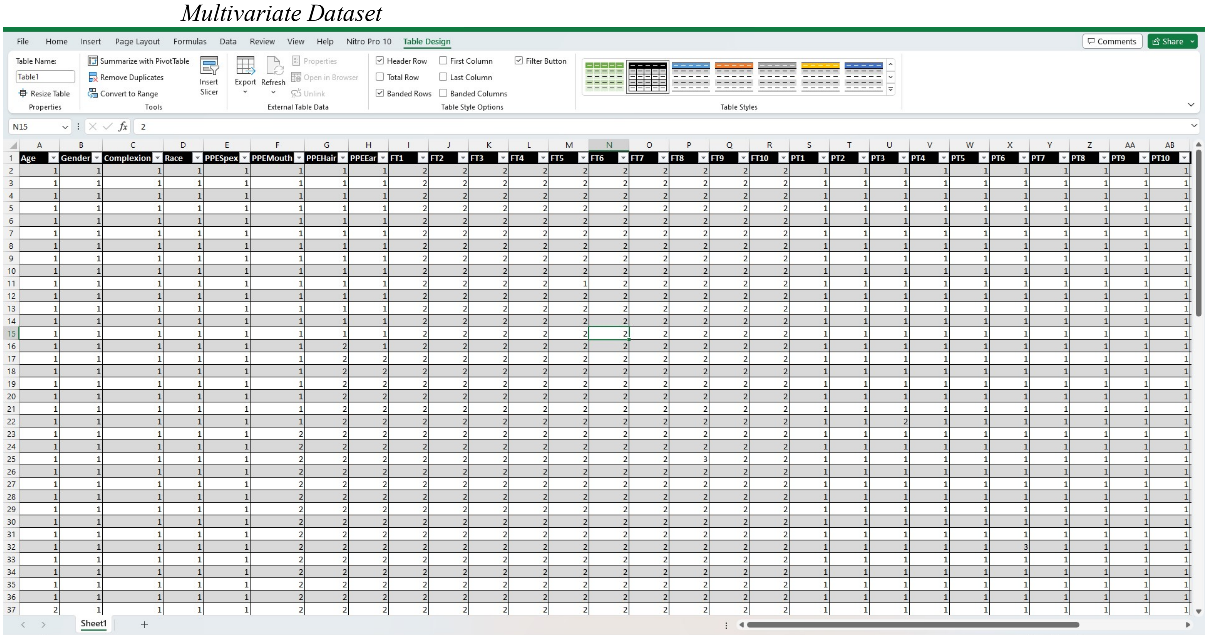


Fig.6(d). PPE-Ear-based Distribution of the Multivariate Dataset

As shown in Fig.6(a),6(b),6(c),6(d), PPE-based objects like Specs were worn by 59% (YES) of the respondents, whereas other respondents were not found in a high percentage with PPE objects as Mouth Mask with 79% (NO), Hair Mask with 86% (NO), Ear with 84% (NO). Hence, it is evident that the demographic analytics of the samples were found to be convincing for the test. The overall outcome of the dataset was created in Excel as shown in Fig.7.



1	Age	Gender	Complexion	Race	PPESpex	PPEMouth	PPEHair	PPEEar	FT1	FT2	FT3	FT4	FT5	FT6	FT7	FT8	FT9	FT10	PT1	PT2	PT3	PT4	PT5	PT6	PT7	PT8	PT9	PT10
2	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1	2	2	2	2	1	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
14	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
15	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
16	1	1	1	1	1	1	2	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
17	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
18	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
19	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
20	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
21	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
22	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	1	1	2	1	1	1	1	1	1	1
23	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
24	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
25	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	3	2	2	1	1	1	1	1	1	1	1	1	1
26	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
27	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
28	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
29	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
30	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
31	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
32	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	3	1	1	1	1
33	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
34	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
35	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
36	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
37	2	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1

Fig.7. Overall compilation of the Multivariate Dataset to test Face detection, Proximity Detection and Object detection in one space

As shown in Fig.7., various distributions of the Multivariate data in the form of Demographics as well as the Visual Assistance Expert Engine Analysis were created using the three major technologies of Computer Vision called Facial Detection, Proximity Detection and Object detection components. These technologies have many different uses and effects. It can help with finding directions and avoiding obstacles, as well as providing more information during conversations. People who can't see well can use technology to

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

help them get around in public places, recognize people's faces, and avoid bumping into things. This makes them feel more confident and can-do things faster. The continuous improvements in these technologies can make them even more precise, versatile, and easily used in our daily lives. This will help create a society where everyone feels included.

5. CONCLUSION

In the future, object and face detection technology to help visually Impaired People look good. In the future, it will work on making detection algorithms better at finding things and faster at doing so. Sophisticated machine learning methods and improved algorithms will help us better recognize and distinguish objects and faces, reducing mistakes and errors. The smooth combination of these technologies into devices like smartphones, smart glasses, or wearable devices will give visually impaired people more accessibility while it out and about. These devices will give instant feedback to help people navigate, recognize objects, and interact with others. In the future, the focus will be on combining different ways of sensing things, such as hearing, feeling, and seeing, to better understand the environment. This approach will improve the user's experience by showing information in a way that is tailored to them and easy to understand. One of the limitations of this model is that it required complex architecture to build real-time systems. AI models will get better at learning and changing in real-time, adjusting to what users like and what's happening around them. These adjustable models will keep getting better at detecting things, making sure that will provide accurate and helpful assistance to people who have trouble seeing. Working together, researchers, developers, advocates for accessibility, and pupils with visual impairments will create solutions that can be used by everyone. Creating universal accessibility standards will guarantee that these technologies are suitable for a wide average of users, promoting inclusiveness worldwide. Combining object and facial detection technologies with smart environments like homes, transportation systems, and public spaces, will make things easier and more convenient for people. Smart spaces will have sensors and connected systems to give help and show directions in real-time.

In summary, object and face detection technologies keep improving and can help visually Impaired people in big ways by making their lives better. These technologies, when carefully created and smoothly combined, will give individuals the power to explore the world with more freedom and self-assurance. By accepting and encouraging these improvements and working together, it can work towards creating a future that is open and welcoming to everyone.

REFERENCES:

[1] Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. Proceedings of the IEEE.

[2]. Naga, P., Marri, S. D., & Borreo, R. (2023). Facial emotion recognition methods, datasets and technologies: A literature survey. Materials Today: Proceedings, 80, 2824-2828.

[3]. Hangaragi, S., Singh, T., & Neelima, N. (2023). Face detection and Recognition using Face Mesh and deep neural network. Procedia Computer Science, 218, 741-749.

[4]. Arkin, E., Yadikar, N., Xu, X., Aysa, A., & Ubul, K. (2023). A survey: Object detection methods from CNN to transformer. Multimedia Tools and Applications, 82(14), 21353-21383.

[5]. Sathana, V., Sneka, S., Sruthika, I., Sujitha, S., & Yogaasri, T. (2023, April). A Soundbite-Based Framework for Text and Object Detection Using OCR and YOLO Technique to Assist Blind and Deaf. In 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 1596-1602). IEEE.

[6]. Choi, J. D., & Kim, M. Y. (2023). A sensor fusion system with thermal infrared camera and LiDAR for autonomous vehicles and deep learning-based object detection. ICT Express, 9(2), 222-227.

[7]. Kang, C. H., & Kim, S. Y. (2023). Real-time object detection and segmentation technology: an analysis of the YOLO algorithm. JMST Advances, 1-8.

[8]. Hemanth, K., & Nagalakshmi, T. J. (2023, April). Improving Accuracy of Face Detection in ID Proofs using CNN and Comparing with DLNN. In 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM) (pp. 1-6). IEEE.

[9]. Arkin, E., Yadikar, N., Xu, X., Aysa, A., & Ubul, K. (2023). A survey: Object detection methods from CNN to transformer. Multimedia Tools and Applications, 82(14), 21353-21383.

[10]. Naik, M. N., Kaur, A., Yathiraju, N., Das, S., & Pant, K. (2023, May). Improved and Accurate Face Mask Detection Using Machine Learning in the Crowded Places. In 2023 3rd

International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 572-576). IEEE.

[11]. Pushpalatha, M. N., Meherishi, H., Vaishnav, A., Anurag Pillai, R., & Gupta, A. (2023). Facial emotion recognition and encoding application for the visually impaired. Neural Computing and Applications, 35(1), 749-755.

[12]. Ezhilarasi, P., Rajagopalan, V. G., Krishnamithran, S., Ramakrishnan, H., & Balaji, H. K. (2023, February). Integrated AI Based Smart Wearable Assistive Device for Visually and Hearing-Impaired People. In 2023 International Conference on Recent Trends in Electronics and Communication (ICRTEC) (pp. 1-6). IEEE.

[13]. Wang, L., Zhang, X., Song, Z., Bi, J., Zhang, G., Wei, H., ... & Zhao, L. (2023). Multi-modal 3D Object Detection in Autonomous Driving: A Survey and Taxonomy. IEEE Transactions on Intelligent Vehicles.

[14]. Giakoumoglou, N., Pechlivani, E. M., & Tzovaras, D. (2023). Generate-Paste-Blend-Detect: Synthetic dataset for object detection in the agriculture domain. Smart Agricultural Technology, 5, 100258.

[15]. Kaur, J., & Singh, W. (2022). Tools, techniques, datasets and application areas for object detection in an image: a review. Multimedia Tools and Applications, 81(27), 38297-38351.

[16]. Diwan, T., Anirudh, G., & Tembhurne, J. V. (2023). Object detection using YOLO: Challenges, architectural successors, datasets and applications. multimedia Tools and Applications, 82(6), 9243-9275.

[17]. Kang, J., Tariq, S., Oh, H., & Woo, S. S. (2022). A survey of deep learning-based object detection methods and datasets for overhead imagery. IEEE Access, 10, 20118-20134.

[18]. Li, K., Chen, K., Wang, H., Hong, L., Ye, C., Han, J., ... & Xu, H. (2022, October). Coda: A real-world road corner case dataset for object detection in autonomous driving. In European Conference on Computer Vision (pp. 406-423). Cham: Springer Nature Switzerland.

[19]. Duan, R., Deng, H., Tian, M., Deng, Y., & Lin, J. (2022). SODA: A large-scale open site object detection dataset for deep learning in construction. Automation in Construction, 142, 104499.

[20]. Kim, J. H., Kim, N., Park, Y. W., & Won, C. S. (2022). Object detection and classification based on YOLO-V5 with improved maritime dataset. Journal of Marine Science and Engineering, 10(3), 377.

[21]. Minderer, M., Gritsenko, A., Stone, A., Neumann, M., Weissenborn, D., Dosovitskiy, A., ... & Houlsby, N. (2022, October). Simple open-vocabulary object detection. In European Conference on Computer Vision (pp. 728-755). Cham: Springer Nature Switzerland.

[22]. Ingle, P. Y., & Kim, Y. G. (2022). Real-time abnormal object detection for video surveillance in smart cities. Sensors, 22(10), 3862.

[23]. Amudhan, A. N., & Sudheer, A. P. (2022). Lightweight and computationally faster Hypermetropic Convolutional Neural Network for small size object detection. Image and Vision Computing, 119, 104396.

[24]. Song, K., Wang, J., Bao, Y., Huang, L., & Yan, Y. (2022). A novel visible-depth-thermal image dataset of salient object detection for robotic visual perception. IEEE/ASME Transactions on Mechatronics.

[25]. Boyd, A., Tinsley, P., Bowyer, K. W., & Czajka, A. (2023). Cyborg: Blending human saliency into the loss improves deep learning-based synthetic face detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 6108-6117).

[26]. Kiruthika, S., & Masilamani, V. (2023). Image quality assessment based fake face detection. Multimed. Tools Appl, 82, 8691-8708.

[27]. Naga, P., Marri, S. D., & Borreo, R. (2023). Facial emotion recognition methods, datasets and technologies: A literature survey. Materials Today: Proceedings, 80, 2824-2828.

[28]. Wang, Z., Huang, B., Wang, G., Yi, P., & Jiang, K. (2023). Masked face recognition dataset and application. IEEE Transactions on Biometrics, Behavior, and Identity Science.

[29]. Hangaragi, S., Singh, T., & Neelima, N. (2023). Face detection and Recognition using Face Mesh and deep neural network. Procedia Computer Science, 218, 741-749.

[30]. Jarraya, I., BenSaid, F., Ouarda, W., Pal, U., & Alimi, A. M. (2023). A new convolutional neural network based on a sparse convolutional layer for animal face detection. Multimedia Tools and Applications, 82(1), 91-124.

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

[31]. Liang, B., Wang, Z., Huang, B., Zou, Q., Wang, Q., & Liang, J. (2023). Depth map guided triplet network for deepfake face detection. Neural Networks, 159, 34-42.

32. Vijaya Kumar, D. T. T., & Mahammad Shafi, R. (2023). A fast feature selection technique for real-time face detection using hybrid optimized region based convolutional neural network. Multimedia Tools and Applications, 82(9), 13719-13732.

[33]. Sheikh, B. U. H., & Zafar, A. (2023). RRFMDS: Rapid real-time face mask detection system for effective COVID-19 monitoring. SN Computer Science, 4(3), 288.

[34]. Firdaus, T. M., Kusumasari, T. F., Suakanto, S., & Pratiwi, O. N. (2023, February). In-House Facial Data Collection for Face Recognition Accuracy: Dataset Accuracy Analysis. In 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE) (pp. 35-39). IEEE.

[35]. Luo, S., Li, X., & Zhang, X. (2023). Wide aspect ratio matching for robust face detection. Multimedia tools and applications, 82(7), 10535-10552.

[36]. Liu, W., Hasan, I., & Liao, S. (2023). Center and scale prediction: Anchor-free approach for pedestrian and face detection. Pattern Recognition, 135, 109071.