ISSN No:-2456-2165

Smart Intelligent Fashion Recommendation System

[1]Hansana A.T.L
Department of Information Technology
Sri Lanka Institute of Information Technology Malabe,
Sri Lanka

[2]Karandawela S.L
Department of Information Technology
Sri Lanka Institute of Information Technology Malabe,
Sri Lanka

[3]Kavindi N.A.H
Department of Information Technology
Sri Lanka Institute of Information TechnologyMalabe,
Sri Lanka

[4]De Silva T.H.H
Department of Information Technology
Sri Lanka Institute of Information Technology Malabe,
Sri Lanka

[5]Dr Lakmini Abeywardena
Department of Information Technology
Sri Lanka Institute of Information TechnologyMalabe, Sri Lanka

Abstract:- This research paper explores the impact of fashion on people's lives and the challenges of online fashion platforms. With only a few people having a clear understanding of fashion suitable for them, online fashion platforms can pose challenges for those less confident in their fashion sense, detracting from the overall shopping experience as the cost of hiring a fashion designer may prove prohibitive for many. The purpose of this research was to provide a solution for finding the perfect matching outfit for people's preferences. Providing the solution to this problem required the critical factors of knowledge about fashion, identifying human body characteristics, gathering the user outfit ideas, recommending a suitable outfit base on the user's ideas, and providing a way to visualize the outcome without FitOn and how to customize the outfit without physically making or buying. This research uses the Smart Intelligence Fashion Recommendation System (SIFR) to address these factors. This system has four components that work together to provide an out- come. Facial expression base intelligent voice assistant and smart mirror component identify the end user's body characteristics component, recommendation component, and human 3D Model creation and fashion customization component. This research pa- per discusses using computer vision, speech recognition, Natural Language Processing, Knowledge Management, recommendation algorithms,3D Model building, hardware resources management and machine learning, and deep learning to build a humanoid Intelligence System.

Keywords:- *Computer Vision, Recommendation Algorithms, Speech-to-Text,3D Model, Natural Language Processing, Knowl- Edge Management, Deep Learning, Machine Learning, Facial Ex- Pression Detection*

I. INTRODUCTION

Fashion and apparel have an enormous impact on people's lives. Because of the quick rate of development, fashion trends frequently alter. In addition, several factors impact modern society, including culture, location, and socioeconomic status, despite personalized shopping experiences. Functioning akin to a virtual mirror, SIFR utilizes captured 2D images of the user to generate a 3D model using deep learning algorithms. This virtual representation is then employed to suggest the best-matching clothing items based on the user's fashion history, body measurements, event type, and skin color. The system's unique feature lies in its human emotion detection component, which identifies the user's facial expressions to determine their level of satisfaction with the recommended fashion patterns. If users are dissatisfied, they may manually rerun the recommendation process or customize their fashion choices. Overall, SIFR represents a state-of-the-art solution that promises to enhance the shopping experience, delivering tailored and intelligent fashion recommendations to users.

II. LITERATURE REVIEW

Previously it has used a multimodal fashion chatbot with enhanced domain expertise. It uses an end-to-end neural con- versational model to create replies based on a categorization- based learning module to capture the fine-grained meanings in pictures. on the conversational past, visual semantics, and subject-matter expertise. Deep reinforcement learning fur- ther improves the model and prevents inconsistent conversa- tion[9]. In human facial expression detection, according to this study[10], gender recognition was faster than the ability to discern between fear and disgust when compared to neutral expressions up to 40u of eccentricity. The visual system could recognize emotional facial expressions peripherally, withterrified faces being recognized more accurately than con-temptuous ones. At 40u of eccentricity, the ability to discern emotion remained above the level of chance. The results imply that the peripheral retina can still grasp emotionally meaningful information essential to social cognition despite the reduced visual resolution in the far peripheral visual field. Microsoft Kinect and augmented reality technology are usedin a dynamic fitting room that allows users to see a live image of themselves as they try on various digital outfits. Using two Kinect cameras, one for obtaining a front image and the other for taking a side image, the system calculates the user's body height based on the head/foot joints and the depth information. The predicted size is close to the users' claimed sizes, according to the evaluation of the proposed model. However, to estimate the measures for this investigation, advanced technology (such as Kinect cameras) is needed[5]. The goals of the suggested method are automated feature point extraction and size measurement on 3D human bodies. The feature extraction and measurement estimation stage is a preprocessing step for virtual fitting or garment designer appli- cations. The suggested approach is automatic and data-driven, unaffected by 3D human body positions and varied shapes. Additionally, the method calls for a depth camera to assess body measurements, which needs to be more appropriate and simple for online buyers to utilize[6]. For building a 3d model using 2D images, it has previously used converting 2D CNN generators to 3D voxel grids.[1][2]. In addition, some research papers proposed building a 3D model based on the point cloudmethod. This unstructured point cloud can generate a 3D mesh with three-dimensional triangles connected by their common edges or vertices.[3]Some research shows that a neural-fields- based IDT technique called Deep Continuous Artefact-free RI Field (DeCAF) can be used to learn a high-quality continuous representation of a RI volume from its intensity-only and limited-angle observations[4]. In the recommendation engine, some research suggested a retrieval method for online fashion images. Finding images of people wearing clothes among images of clothing can be rather challenging. In this situation, traditional image retrieval techniques could be more useful. The four sections of the full-body fashion coordinate image aredivided, and an image is returned with the target area's equiv- alent clothing image[7].In addition, some researchers used a recurrent neural network (RNN) and a dynamic, collaborative filtering technique to build a recommendation system. The RNN-based recommendation system examines each person's preferred styles based on anything from a single purchase price to a string of sales occurrences. A popularity ranking baseline strategy had an Area under the Curve value of 80.2 percent, whereas the proposed Recurrent Neural Network model had a value of 88.5 percent[8]. However, in the proposed system, we focus on centralizing every component mentioned above in a single hardware unit to utilize the fashion recommendation.

III. METHODOLOGY

A. The "Smart, Intelligent Fashion Recommendation System"Solution is Mainly Consisting of Four Main Components,

- ➢ *Facial Expression Base Intelligent Voice Assistant*
- ➢ *Body Measurement Calculation System*
- ➢ *Human 3D Model Creation and Fashion Customization*
- ➢ *Fashion Recommendation*

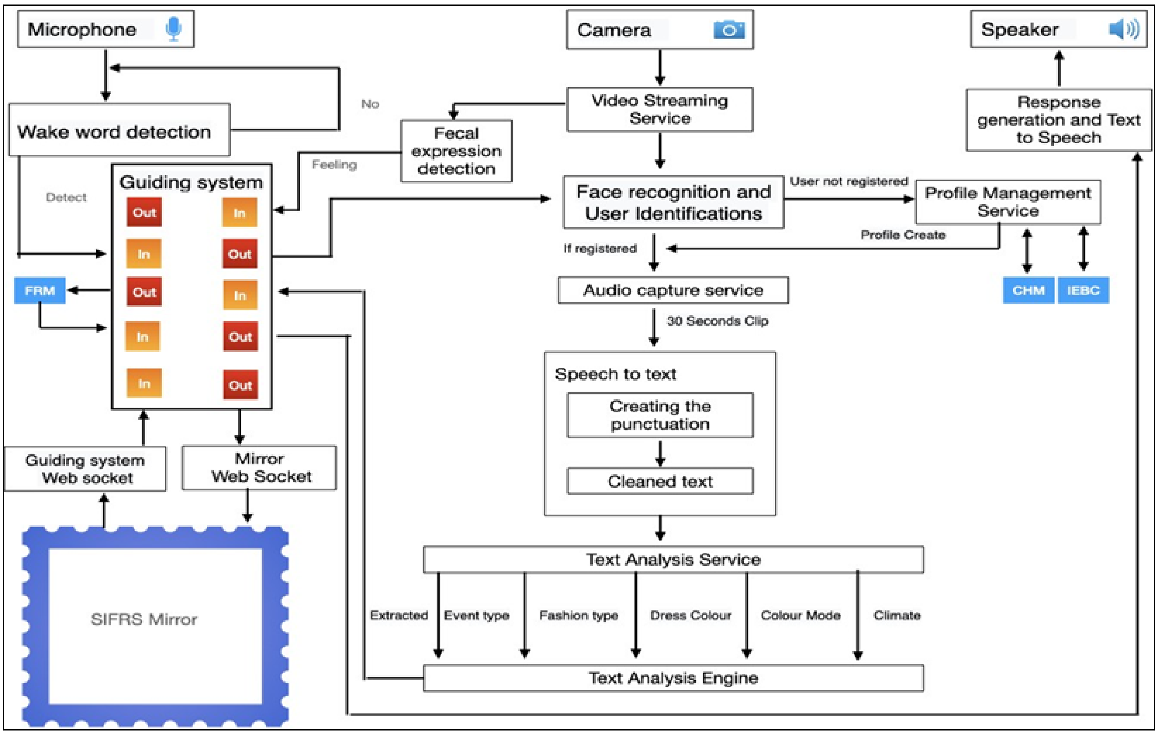


Fig 1 Overall System Diagram

- *Facial Expression base Intelligent Voice Assistant*

The facial expression-based intelligent voice assistant (FEBIVA) is an advanced system that can understand the user's words and provide solutions based on that information. It can also detect and react to users' emotional states to improve communication. The FEBIVA system uses technologies like emotion recognition, voice recognition, and knowledge-based problem-solving to understand user needs and preferences. By learning from user interactions and feedback, FEBIVA can expand its knowledge and provide better services to its users. This system focuses on leveraging natural human behavior to enhance the capabilities of computer systems.

- *Facial Expressions Detection:*

It uses Computer vision and deep learning models to recognize and analyze user facial expressions. By examining the position and movement of facial parts such as eyebrows, mouth, and eyes, this system accurately determines the user's emotions, such as happi- ness, neutrality, and sadness. The component connects to a video stream service and reads each frame to identify frames containing human faces. Then, using the open CV cascade classifier frontal face, it extracts data from the detected human faces. The extracted region is then resized to 244x244 pixels, normalized, and converted into a NumPy array. This array is then passed to the model for prediction and relayed to the Guiding System. The MobileNet model is a lightweight deep convolutional neural network (CNN) model. The training data used was the Kaggle FER2013 dataset. The image size is 48x48 pixels, but the MobileNet model requires imagesof 224x224 pixels. Write the Python script to automate this resizing process for all images in the Kaggle FER2013 dataset.

- *Wake Word Detection:*

Keeps the voice assistant off untila specific keyword is recognized, such as "Sifer," "Hi Sifer," or "Hello Sifer." The wake word recognition process consists of two main components: the listener and the wake word engine. The listener adds a 2-second Wave audio clip to the audio queue. This audio clip sample rate is 8000. Wake Word Engine takes each item from the queue, converts the wave file to MFCC, and creates the PyTorch tensor. These tensors are fed into a Wake word deep learning model to perform binary classification to determine whether an audio clip contains wake words. If the wake word is detected, the model returns a value of 1 and is not return 0. The dataset consists of two main categories of wave file folders. A folder named Folder Name 1 contains audio clips where people say the wake word "SIFER"The duration of these clips is 2 seconds, and the samplingrate is 8000. Each clip name is a unique recording index,such as 1.wav and 2. wav. The second Folder (folder name 0) contains audio clips without wake-up words. This Folder also has background noise. The non-wake word clips in this Folder clips get from the Mozilla Common Voice dataset, specificallyCommon Voice Delta Segment 12.0, and are in mp3 format. The preprocessing script converts these mp3 files into wave files. Then each audio clip is divided into 2 seconds. Label 0 represents 17,238 non-wake word wave files, while label one only has 110 wake word clips. Up-sampling techniques are used to generate more data and mitigate the inequality in the dataset.

- *Speech-to-Text :*

System has two main components: the audio recording service and the conversion of the recorded speech to text. The PyAudio library allows accessing the microphone and reading audio data at a sample rate of 16000 using the Mon channel configuration. Audio data is read in 16-bit format in 1024 chunks, and each chunk is added to a queue. Each audio clip can be up to 30 seconds long, andthe queue contains a byte array representing 1 second of audio data. The next step is to convert the speech data to text using the Vosk Speech-to-Text Toolkit. Vosk is an open-sourcetoolkit known for its accurate and efficient speech recognition capabilities. One of its main advantages is its user-friendly nature, which provides a simple API for integrating speech recognition functionality. However, the output produced by Vosk does not contain punctuation. Instead, it uses a recursion and punctuation model based on Bert. Finally, the obtained text is sent to the text analysis component using Apache Kafka to free up microphone resources.

- *Text Analysis:*

The content discusses text analysis of consumers and their responsibilities. The primary responsi- bility of consumers is to listen to an Apache Kafka topic and capture cleaned text. There are two services for text analysis consumers. The first service handles Kafka-related tasks such as consuming messages and producing extracted data for other services. The second service is the actual text analysis. It performs two tasks: encoding the text using the Bert tokenizer encode plus method and loading the relevant text analysis model. The preprocessing value is passed to the model's prediction method to obtain a prediction value probability array using the NumPy max method. The arg max function is used to find the index of the highest probability, which corresponds to the predicted label. The text analysis engine receives this data, monitors the number of runningconsumers, and builds the recommendation model based on the received payloads. Once the engine receives multiple payloads from different consumers, it sends the final parameter payloadto the guiding system using Apache Kafka. The consumers send a heartbeat signal to the text analysis engine, allowing it to count the number of active consumers.

- *Face Recognition and user Identification:*

Usually use deep learning algorithms to identify, recognize and match faces. These systems work by extracting facial features or embeddings from images and comparing them to databases of known faces. Here's an overview of how the process works:

- *Response Generation :*

This component generates a re- sponse based on the user's request, considering the informationgathered from the other components. The response can be a spoken message.

- *Guiding System :*

Of the brain is the SIFER system, which has two leading roles. The first is monitoring all system events, resource usage, and input/output (IO). The second responsibility is communication between components. The guiding system divides the main tree into three sub-components (decision unit, memory unit, flow control unit), each of which plays a specific role.

- *Mirror Application:*

Mirror Applications is the Font-end in the visual representation of this system. This Application wrote in the Dart and Flutter Framework. This Application maintains multiple states to provide a reactive user experience. That is choosing the flutter for the build of this Application. Another advantage is cross-platform's Single code base can use multiple platforms. This mirror application runs the Mac OS Linux, Androids, and Windows platforms in further enhancement if required to move to another platform—no need to rewrite this Application.

➢ *Body Measurement Calculation System*

MediaPipe is an open-source framework for creating image and video processing pipelines. It uses advanced computer vision algorithms to accurately determine key body points such as elbows, wrists, shoulders, and hips. This represents a significant advance in anthropometric calculation systems. Mathematical equations and algorithms are applied through MediaPipe to determine body measurements such as bust, waist, hip, and inseam. Accurate and immediate body mea- surement depends on the accuracy of detected body points. The resulting pixel measurements must be converted to metric measurements using a scale factor. It allows the practical use of body measurements in units such as centimeters and inches, which is helpful for clothing designers and manufacturers. Thefinal step of the anthropometric calculation system involves creating a size chart. This chart converts body sizes to standardsizes such as S, M, and L to give the user an accurate fit and comfort. Skin color detection uses computer vision technology to detect human skin color. By taking an imageand processing it using OpenCV, the Haar Cascade Classifier separates faces from the rest of the image. A skin color filteris then applied to separate skin pixels based on a predefined range in the YCrCb color space. Denoising techniques suchas morphological operations and filtering are used to enhance the detection of skin pixels. The hexadecimal code of the skin color is obtained by analyzing the average RGB value of the skin pixels. To calculate the melanin index, the pixels of the same skin color are converted to the LAB color space. The LAB color space classifies human skin types based on melanin content and response to sunlight exposure. This classification system is essential for personalized skin care recommendationsand melanoma screening.

➢ *Human 3D Model Creation and Fashion Customization*

Our approach involves using a modified deep learning model trained on a dataset of human images to generate a 3D model that can be used for virtual try-on applications. We first preprocess the 2D images to extract silhouettes, surface normals, and camera parameters. The silhouettes are binary masks that indicate the boundary of the human bodyin the image. The surface normals indicate the orientation of the human body's surface at each pixel. Finally, the camera parameters specify the position and orientation of the camera that captured the image. We then train a modified deep learning model on the dataset of human images to generate a 3D model that can be used for virtual try-on applications. Our modified model includes additional layers to generate a more detailed 3D human body model. We use a combination of convolutional neural networks (CNNs) and fully connected neural networks (FCNs) to generate the 3D model. After generating the 3D model, we add clothes to the generated 3D model to create a complete 3D representation of the human body. First, weuse Blender, a 3D modeling software, to add clothes to the generated 3D model. Next, we import the generated 3D modelinto Blender and use Blender's GUI or Python API to add clothing items and attach them to the model. We then export the complete 3D model as a file that can be used for virtual try-on applications.

➢ *Fashion Recommendation System*

A fashion recommendation model analyzes a user's prefer- ences and suggests clothing that fits their body type and style. This algorithm collects data such as age group, gender, skin color, fashion preferences, and types of events the user wears. It also considers the user's clothing size, body measurements, preferred colors, and the weather conditions of the venue. The system can rank clothes based on user preferences and budget factors. This model uses a collaborative filtering algorithm called item-based collaborative filtering that matches user input with fashion items in a database. Preprocess the dataset and calculate similarity scores between items using cosine similarity to identify similar items based on user behavior. In addition, a content-based approach is used to collect in- formation about the characteristics of fashion items, such as color and style. This approach uses a decision tree algorithm totrain fashion item features to recommend items based on user preferences. If the user is satisfied with the recommendation,it is saved as a positive result. If a user still needs to, theuser can re-enact the recommendations or customize the dress patterns based on the user's feelings.

IV. RESULTS

A. *Body Measurement Calculation System*

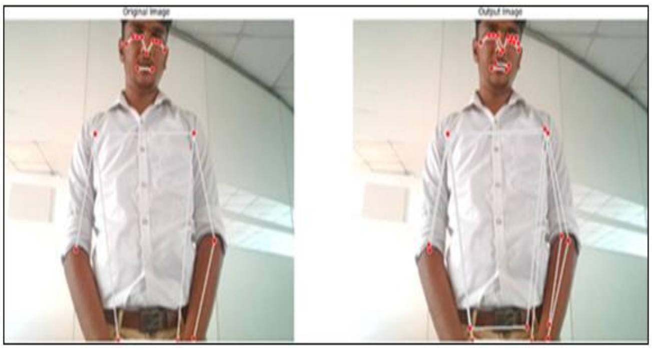
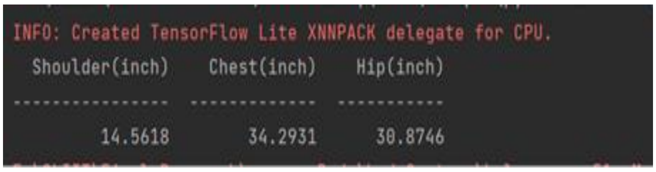


Fig 2 Body Measurement Calculation



```
INFO: Created TensorFlow Lite XNNPACK delegate for CPU.
  Shoulder(inch)    Chest(inch)    Hip(inch)
----------------  -------------  -----------
         14.5618        34.2931      30.8746
```

Fig 3 Body Measurement Values

After using a media pipe, the body measurement calculationsystem can identify the human body pointers, and using relevant body indexes, it is used to determine x and y axises points of body indexes. Using the calculated measurements, we show the predicted size findings. We assessed the clothing size and compared the estimated size to the actual size of the participants' clothing. We tested the sample of 42 participants, 25 of whom were men and 17 of whom were women. Weused shoulder width and chest size as features to anticipate the upper body measurements for the first group. Results predicted upper body measurements demonstrate that 14 out of 25 males were expected to have the same genuine size, showing the components' accuracy. For group 2, 9 of 17 females have the same clothing size.

B. Skin Color Identification System

After extracting the skin color of the previously mentioned group of people, out of 25 males, we received 16 with the same skin Fitzpatrick type, and from the second group, wehad 12 out of 17 having the same skin type.

C. Facial Expression base Intelligent Voice Assistant

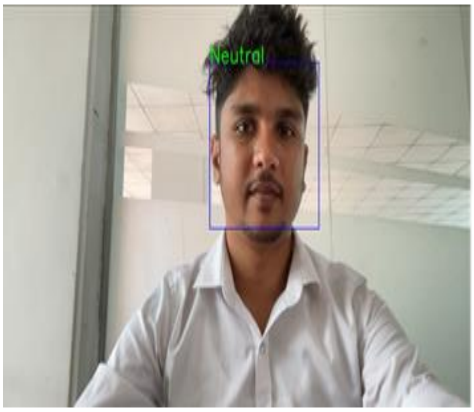


Fig 4 Neutral Facial Expression

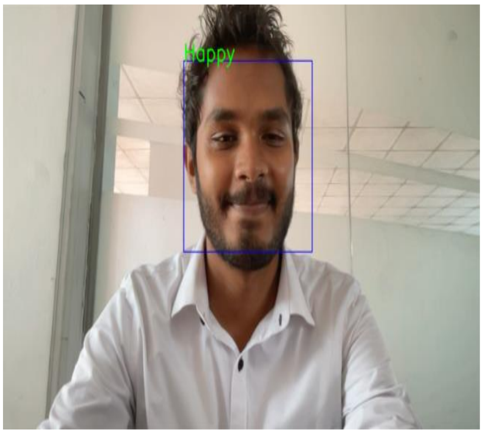
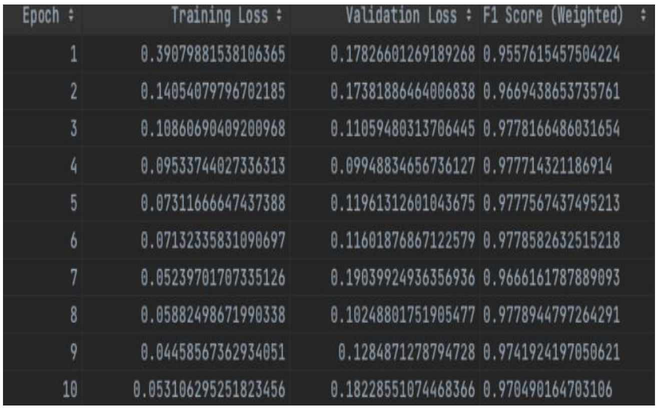
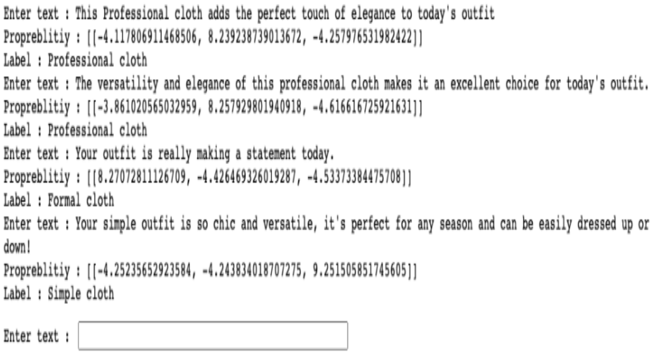


Fig 5 Happy Facial Expression



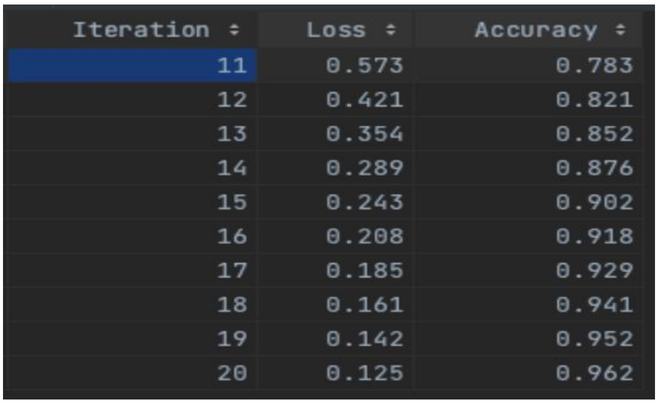
Epoch ⇕	Training Loss ⇕	Validation Loss ⇕	F1 Score (Weighted) ⇕
1	0.39079881538106365	0.17826601269189268	0.9557615457504224
2	0.14054079796702185	0.17381886464006838	0.9669438653735761
3	0.10860690409200968	0.11059480313706445	0.9778166486031654
4	0.09533744027336313	0.09948834656736127	0.977714321186914
5	0.07311666647437388	0.11961312601043675	0.9777567437495213
6	0.07132335831090697	0.11601876867122579	0.9778582632515218
7	0.05239701707335126	0.19039924936356936	0.9666161787889093
8	0.05882498671990338	0.10248801751905477	0.9778944797264291
9	0.04458567362934051	0.1284871278794728	0.9741924197050621
10	0.053106295251823456	0.18228551074468366	0.970490164703106

Fig 6 Accuracy of the Facial Expression Model



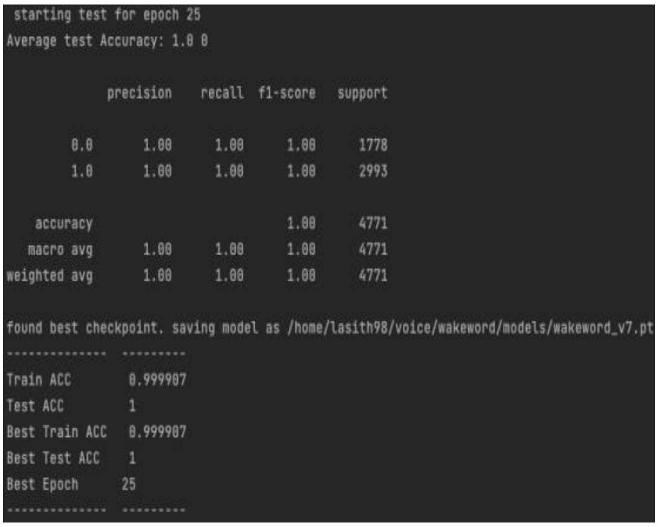
```
Enter text : This Professional cloth adds the perfect touch of elegance to today's outfit
Propreblitiy : [[-4.117806911468506, 8.239238739013672, -4.257976531982422]]
Label : Professional cloth
Enter text : The versatility and elegance of this professional cloth makes it an excellent choice for today's outfit.
Propreblitiy : [[-3.861020565032959, 8.257929801940918, -4.616616725921631]]
Label : Professional cloth
Enter text : Your outfit is really making a statement today.
Propreblitiy : [[8.27072811126709, -4.426469326019287, -4.53373384475708]]
Label : Formal cloth
Enter text : Your simple outfit is so chic and versatile, it's perfect for any season and can be easily dressed up or
down!
Propreblitiy : [[-4.25235652923584, -4.243834018707275, 9.251505851745605]]
Label : Simple cloth

Enter text :
```

Fig 7 Text Analysis



Iteration ⇕	Loss ⇕	Accuracy ⇕
11	0.573	0.783
12	0.421	0.821
13	0.354	0.852
14	0.289	0.876
15	0.243	0.902
16	0.208	0.918
17	0.185	0.929
18	0.161	0.941
19	0.142	0.952
20	0.125	0.962

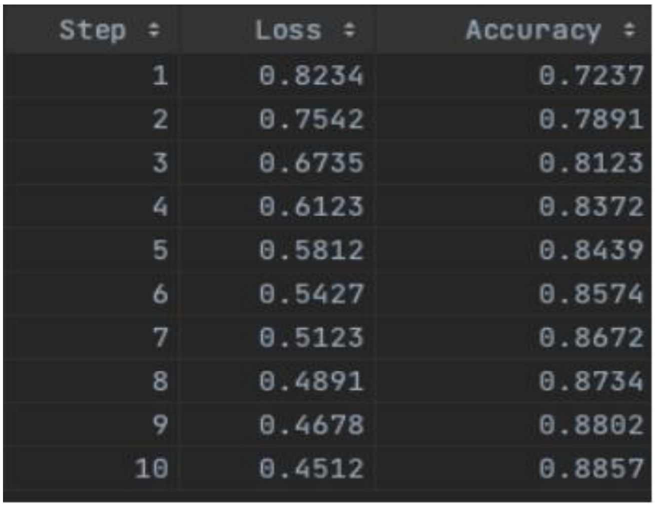
Fig 8 Accuracy of the Text Analysis model



```
 starting test for epoch 25
Average test Accuracy: 1.0 0

              precision    recall  f1-score   support

         0.0       1.00      1.00      1.00      1778
         1.0       1.00      1.00      1.00      2993

    accuracy                           1.00      4771
   macro avg       1.00      1.00      1.00      4771
weighted avg       1.00      1.00      1.00      4771

found best checkpoint. saving model as /home/lasith98/voice/wakeword/models/wakeword_v7.pt
--------------  ---------
Train ACC        0.999907
Test ACC         1
Best Train ACC   0.999907
Best Test ACC    1
Best Epoch      25
--------------  ---------
```

Fig 9 Accuracy of the wake word model

D. 3D Model Build and Fashion Customization System



Step ⇕	Loss ⇕	Accuracy ⇕
1	0.8234	0.7237
2	0.7542	0.7891
3	0.6735	0.8123
4	0.6123	0.8372
5	0.5812	0.8439
6	0.5427	0.8574
7	0.5123	0.8672
8	0.4891	0.8734
9	0.4678	0.8802
10	0.4512	0.8857

Fig 10 Accuracy of the Customization Model

E. Fashion Recommendation System

The fashion Recommendation model that uses CNN models for image processing can extract more intricate visual features and provide more accurate recommendations; the CNN model is used to analyze user preferences and recommend outfits that match their body type, style, and event type.

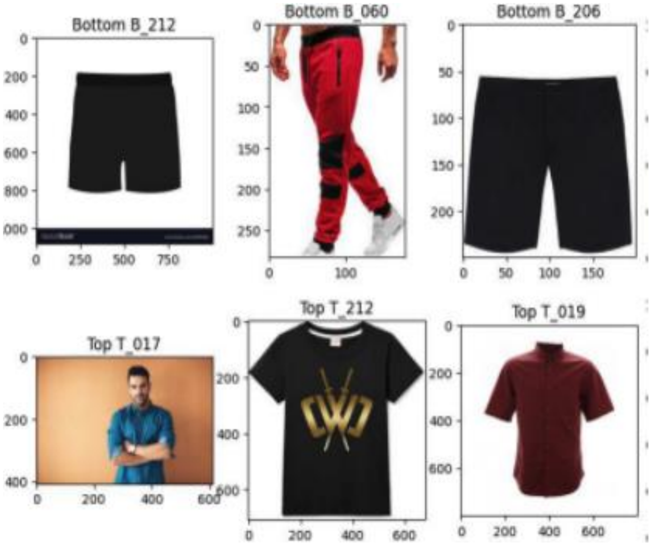


Fig 11 Output Recommendations

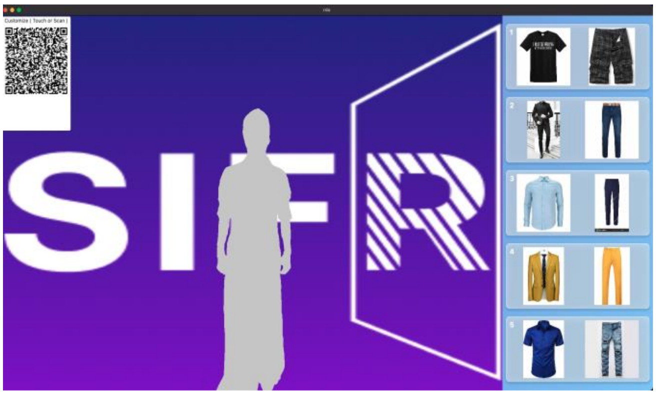


Fig 12 Mirror Application

V. CONCLUSION

This application's primary goal is to provide a hardware system with an enhanced user experience to recommend fashion patterns to users who need more knowledge about fashion. The main problems faced in the fashion industry, like problems identifying body measurements and the problems identifying the perfect clothes matching for the specific user and events, are resolved in this proposed system. In addition, the proposed system can enhance the user experience with an inbuilt intelligent Voice assistant, face emotion identification system, and body characteristics identification system. Instead of recommending the clothes, the proposed system automat- ically gathers the desired inputs using a skin color detection system and body measurement calculation system. In addition,the intelligent Voice assistant will connect all the system's subcomponents and then connect the user and another system more humanly. Most recommend the clothes, but if a system- identified user is not satisfied with the recommendation using a facial emotion recognition system, the user can customize or rerun it. The 3D model component will display the human 3D dummy so users can see their way more efficiently. More- over, the recommendation system suggests the best matching clothes to the user using the input gathered using the body characteristics identification system and voice assistant. The recommendation system helps users get the perfect outfits for specific events and characteristics, such as skin type and body measurements.

REFERENCES

[1]. M. A. A. A. Sahar Ashmawia, FITME: BODY MEASUREMENT ESTIMATIONS USING.

[2]. D. C. B. J. M. K. S. Dewan, Estimate human body measurement from a 2D image using computer vision, 2022.

[3]. Q. T. a. L. Dong, An Intelligent Personalized Fashion Recommendation System, 2010.

[4]. H. P. o. L. Tan Xiao, Automatic human body feature extraction and personal size measurement, 2017.

[5]. K. B. Shaik, P. Ganesan, V. Kalist, B. Sathish and J. M. M. Jenitha, Comparative Study of Skin Color Detection and Segmentation in, 2015.

[6]. K. Liu, J. Wang, E. Kamalha, V. Li and X. Zeng, Construction of a prediction model for body dimensions used in garment pattern making based on anthropometric data learning., 2017.

[7]. P. Meunier, Performance of a 2D image-based anthropometric measurement and clothing sizing system, 2010.

[8]. O. H. Erich Stark, Low-Cost Method for 3D Body Measurement Based on Photogrammetry.

[9]. Y. Chae, J. Xu, B. Stenger and S. Masuko, "Color navigation by qualitative attributes for fashion recommendation," 2018 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 2018, pp. 1-3, doi: 10.1109/ICCE.2018.8326138.

[10]. S. Verma, S. Anand, C. Arora and A. Rai, ”Diversity in Fashion Recommendation Using Semantic Parsing,” 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 2018, pp. 500- 504, doi: 10.1109/ICIP.2018.8451164.

[11]. M. Iso and I. Shimizu, ”Fashion Recommendation System Reflecting Individual’s Preferred Style,” 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE), Kyoto, Japan, 2021, pp. 434-435, doi: 10.1109/GCCE53005.2021.9622080.

[12]. S. O. Mohammadi, H. Bodaghi and A. Kalhor, ”Single-Item Fashion Recommender: Towards Cross-Domain Recommendations,” 2022 30th International Conference on Electrical Engineering (ICEE), Tehran, Iran, Islamic Republic of, 2022, pp. 12-16, doi: 10.1109/ICEE55646.2022.9827421.

[13]. Pereira, Tiago Matta, Arthur Mayea, Carlos Pereira, Frederico Monroy, Nelson Jorge, Joao Rosa, Tiago Salgado, Carlos Lima, A. ˜ Machado, Ricardo-J Magalhaes, Lu ˜ ´ıs Adao, Telmo Guevara Lopez, ˜ Miguel Angel Garcia, Dibet. (2021). A web-based Voice Interaction framework proposal for enhancing Information Systems user experience. Procedia Computer Science. 196. 235-244. 10.1016/j.procs.2021.12.010.

[14]. Anil Audumbar Pise, Mejdal A. Alqahtani, Priti Verma, Purushothama K, Dimitrios A. Karras, Prathibha S, Awal Halifa, ”Methods for Facial Expression Recognition with Applications in Challenging Situations”, Computational Intelligence and Neuroscience, vol. 2022, Article ID 9261438, 17 pages, 2022. https://doi.org/10.1155/2022/9261438

[15]. SusheelKumar,Vijay Bhaskar Semwal,Shitala Prasad, Generating 3D Model Using 2D Images of, 2017.

[16]. B. K. Hashim Yasin, An Efficient 3D Human Pose Retrieval and Reconstruction from, 2018.

[17]. I. Elkhrachy, 3D Structure from 2D Dimensional Images Using Structure, 2022

[18]. K.-Z. G. Swarna Priya, 3D reconstruction of a scene from multiple 2D images, 2017