Health Trauma and Well-being Assistant for Bengali Seniors in Household: A Multimodal Approach

by

Md. Nazmul Islam
19301033
Md. Al-Amin
19301028
Md. Zaed Hassan
19301024
Tamzid Islam
19301050

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
School of Data and Sciences
Brac University
January 2024

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Md.Nazmul Islam 19301033	Md.Al-Amin 19301028
Md.Zaed Hassan 19301024	Tamzid Islam 19301050

Approval

The thesis titled "Health Trauma and Well-being Assistant for Bengali Seniors in Household: A Multimodal Approach" submitted by

1. Md. Nazmul Islam (19301033)
2. Md. Al-Amin (19301028)
3. Md. Zaed Hassan (19301024)
4. Tamzid Islam (19301050)

Of Fall, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on January 18 , 2024.

Supervisor:
(Member)

Dr. Md. Khalilur Rhaman
Professor
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Ethics Statement

In doing research for this thesis, we commit to observe the highest ethical standards by obtaining informed consent from all participants, maintaining the confidentiality of sensitive information, and conducting our work with integrity and neutrality.

Abstract

The increasing number of elderly individuals living alone has emerged as a pressing global concern. Our research aims to address this issue by developing advanced modules that can be integrated into a system that enhances the quality of life for older adults. The modules focus on medicine detection, fall detection, reminders for important tasks and events and providing companionship through friendly verbal interactions. Through the integration of cutting-edge deep learning techniques, diverse models and natural language processing (NLP), we have successfully designed an effective medication and well-being assistant. These modules use computer vision technology along with reinforced learning from human feedback and convolutional neural networks (CNNs) to reach our goal. The modules can be integrated into systems to empower elderly individuals to lead more active and fulfilling lives. Finally, this research contributes to the well-being and happiness of the elderly, highlighting the significance of comprehensive support systems in promoting their overall well-being.

Keywords: elderly people; companion system; medication habit; deep learning; NLP; computer vision; reinforced learning; CNN; medication assistant

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis has been completed without any major interruption.
Secondly, we would like to express our gratitude to our supervisor Dr. Md. Khalilur Rhaman sir, for his kind support and guidelines throughout our research. His teachings and guidelines have helped us inspire most of the ideas expressed in this paper. Additionally, the research assistant of our supervisor Sayantan Roy Arko sir was always available to support us when we needed help. Finally, our parents' support was vital; without their kind support and prayers, our research would not be possible.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

API Application Programming Interface

CNN Convolutional Neural Networks

CSP Constraint Satisfaction Problems

CTT Concur Task Trees

$EOFM$ Enhanced Operator Function

IoU Intersection over Union

$LLMs$ Large Language Models

MAP Mean Average Precision

$NICO$ Neuro-Inspired Companion

NLP Natural Language Processing

NMS Non-Maximum Suppression

OpenCV Open Source Computer Vision

$R-CNN$ Region-Based Convolutional Neural Network

RAG Retrieval Augmented Generation

$SMTP$ Simple Mail Transfer Protocol

SSD Single Shot Detector

SVM Support Vector Machine

$YOLO$ You Only Look Once

Chapter 1

Introduction

In recent times, the number of elderly populations is constantly growing all over the world. The results of a recent survey conducted during the COVID-19 pandemic over two weeks in April 2020 among 500 individuals aged 70 and above raised alarming concerns about the increasing loneliness epidemic during the lockdown. The survey specifically highlights the distressing situation faced by those who live alone. It reveals that over half of the respondents in this group have limited or no contact with their families, and they are three times more likely to experience fears of being left alone. These findings emphasize the urgent need to address the growing issue of loneliness, especially among older individuals living alone, during these challenging times.[12] The nuclear family is the most popular idea nowadays for family planning which leaves the elderly people in isolated houses on their own. With the increasing challenges of elder care, researchers are exploring the use of assistant device services in care centers and households. This has long been a human dream. Our objective is to create modules that will act as a medication and well-being assistant for the elderly. Assistive features offer immense potential to enhance the lives of older adults, enabling them to maintain independence and a sense of control as they age. Extensive global user studies involving thousands of elderly individuals, along with their families and caregivers, have provided valuable insights into their unique needs and preferences for home robot companions. These studies consistently reveal a positive reception towards healthcare robots that support both physical and mental well-being. There are several companion robots available in the market such as Eilikbot, Jibo, Aibo, Misa, etc. But none of them satisfies all the needs of elderly people who lives in Bangladesh. Consequently,we propose a system that will have modules which will be easier to use for the elderly living in Bangladesh and will be effective for their physical and mental well-being. We want to use a deep learning-based model to detect medicines in real-time. We also want to make our system friendly for the elderly with the help of NLP and Reinforced learning. Furthermore, we want to use another model to track the movement of the elderly and monitor their daily activities to keep them safe. Finally, by feeding more data to our system we want to update our system frequently and keep it up to date. Thus, with technology, we want to permanently raise lifestyle quality for elderly people living in Bangladesh. Figure 1.1 shows the widespread social isolation and loneliness.

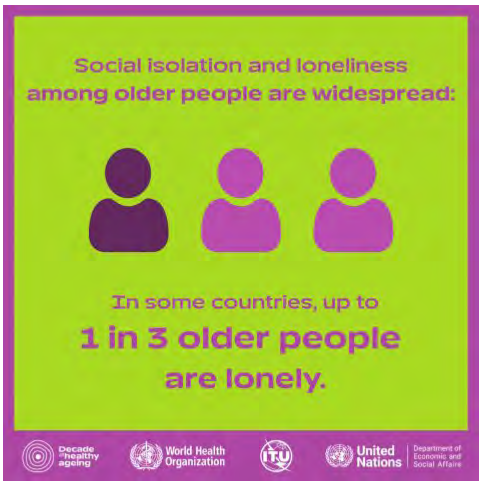


Figure 1.1: Social isolation and loneliness [12] .

1.1 Problem Statement

The problem we are trying to address here is the absence of a decent companion in the lives of the elderly who live alone. From various research done before, we found a companion right now is needed for elderly people more than ever. More than one in three older adults (37%) reported feeling a lack of companionship (29% some of the time, 8% often) in the past year, compared with 41% (32% some of the time, 9% often) in 2020, and 34% (26% some of the time, 8% often) in 2018. [22] Aging comes with a certain loss in abilities such as weak eyesight, low hearing power, weak memory, body numbness, etc. The elderly people who live without caregivers face various kinds of problems due to their weakened bodies. We found that In Taiwan, the aging index exceeded 100 in February 2017. [5] Likewise, the elderly population accounts for at least 27% of the Japanese population. [5] From another study we found that life expectancy in India continues to rise each year, indicating an increase in the average lifespan. The Population Census of 2011 revealed that India has a considerable number of senior citizens, with approximately 104 million individuals aged 60 or above. Unfortunately, many of these elderly people experience social isolation and feelings of loneliness due to the prevalence of nuclear family structures, which often results in reduced social support networks for the elderly. [6] According to the data from the World Bank, life expectancy in India is 68.35 years as of 2015. As per the Population Census of 2011, almost 15 million elderly Indians live all alone (humanoid robots as companions). We also conducted our research to better understand the problems and behaviors of the elderly. We randomly selected 100 elderly people aged between 50 to 100 for our survey. This is the segment we are researching for. Figure 1.2 shows the changes in feelings of social isolation trend. If we leave this problem as it is, many of our elderly will be kept at continuous risk. They may forget important work and important meetings or even forget to take medicines on time. Having our elders at risk can even lead to life-threatening situations as well. Also, we found from various reports that living alone for a long time without any companion has severe harmful effects on the human body and mind. Nowadays, the number of elderly people living alone in the home is increasing every year. However, the number of trained caregivers is not enough. Elderly people need to be aware of their medicines, their sleep time, work schedules, and prayer

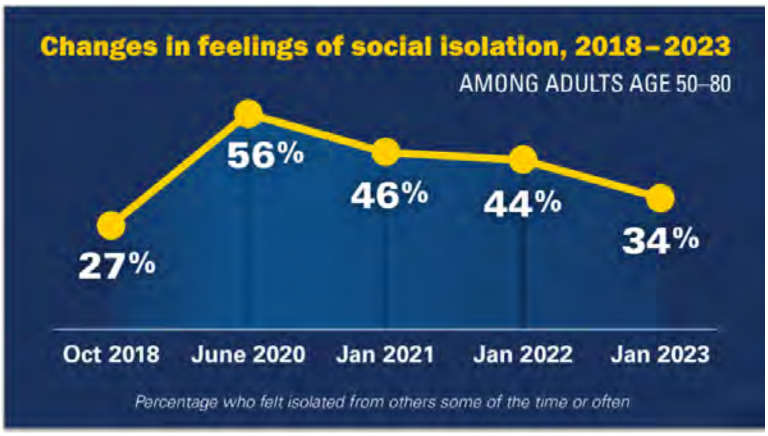


Figure 1.2: Changes in Feelings of Social isolation trend [22] .

times and most importantly they need a caregiver who will not let them feel bored. From various research done before, we found that there are several companion bots available in the market, also there are separate systems for reminder and household work. However, there is no system available in the market that will be able to become a personal companion for elderly people in Bangladesh. Figure 1.3 shows the percentage of men and women in terms of feelings of loneliness.

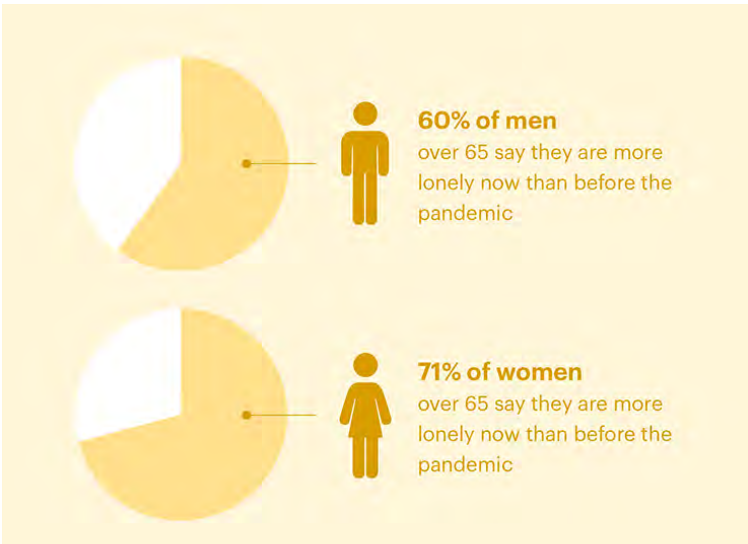


Figure 1.3: Feeling of loneliness between men and women [18] .

1.2 Survey Findings

We surveyed for our research among 100 elderly people all aged between 50 to 100 years. We found some interesting facts regarding their daily behaviors. We found that about 44.7% of our respondents are female while the other 55.3% are male. We also found out the topics they like to talk about every day which are interesting and diversified. The topics are shown in Table 1.1 and Table 1.2.
We also asked them what works and situations they feel challenged for. The response was amazing. We found that the main challenges for the elderly are picking

Politics	Memories	Old Days	Sports	Gossip	বাড়ির আশে পাশে কোন কোন ঘটনা হলো
জীবনমুখী কথা	She is now out of her mind	Daily Life	Sometimes she doesn't know what she is talking about.	জমি সংক্রান্ত	Recent politics
কার কোথায় বিয়ে হলো, কয়টা বাচ্চা হলো	মাজার ভিত্তিক	About politics and values	বাড়ির আশে পাশে কোন কোন ঘটনা হলো	গ্রাম্য পলিটিক্স	নাতি নাতনিদের বিষয় এ
খবর	Religion, Tv series, life	জমি সংক্রান্ত	ওনার চাকুরী জীবন।	কোন এলাকায় কি হয়েছে সেগুলো নিয়ে	মুক্তিযুদ্ধের সময়ের
আগের দিনের কথা	ঠাকুর দেবতা	খুব কম কথা বলেন।	রূপচর্চা	ঘরের কে কি দোষ করলো সেগুলো নিয়ে।	খবর

Table 1.1: Survey Findings 01

টিভি সিরিয়াল	religion	Movies	যে কোনো বই নিয়ে	কর্মজীবন	ধর্ম, তাবলীগ
প্রতিদিন আশেপাশে যা ঘটে তা নিয়ে।	স্টার জলসা, জি বাংলার নাটক, বাংলা ছবি	Farming	আশেপাশে কি ঘটলো সেগুলো নিয়ে।	গ্রাম্য আলাপন	পারিবারিক
সংসার নিয়ে	নিজের ছোট বেলার গল্প	সমসাময়িক ঘটনা	Shangsarik	পুরোনো দিনের কথা	জমিজমা
ধর্মীয় বিষয়	তার যুবক বয়সের গল্প	রাজনীতি	About her grandchildren	স্মৃতিচারণ	বই, পড়াশোনা
যুদ্ধের সময় কি করেছিলেন।	ডিমেনশিয়ার রুগি	বাপ দাদার আমলের কাহিনি	পারিবারিক কথা	Unadr somoy ki chilo ekhn ki ache egla niya	কড়া রাজনীতি
ঘটকালির	Business related	নাতি নাতনি নিয়ে	Judgmental	সংসার নিয়ে	তাবলীগ

Table 1.2: Survey Findings 02

something up, climbing stairs, walking for a substantial amount of time, and taking medicine. Other significant mentions are using washrooms, cooking, using electronic devices, and so on. The findings are shown in Figure 1.4. We also asked them about

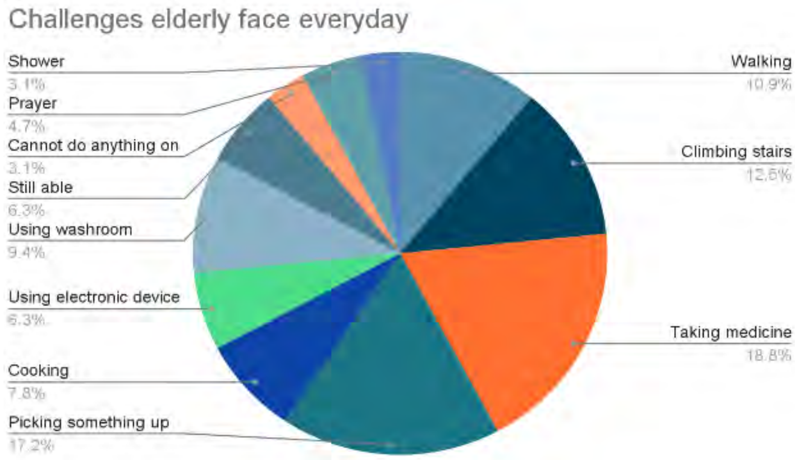


Figure 1.4: Pie chart of challenges faced by elderly

their daily needed items, shown in Figure 1.5. We can see from the chart that most

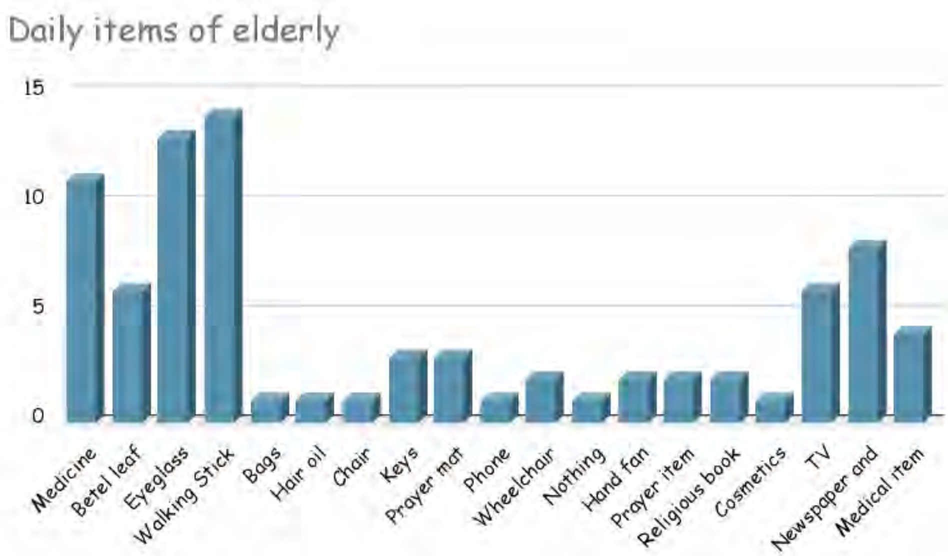


Figure 1.5: Bar chart of item needs of Bengali Senior

of the elderly people need medicine daily. They also need eyeglasses and walking sticks to support themselves. Sometimes they look for TV and Newspapers. A significant amount of the elderly need medical items such as hearing machines, diabetes machines, blood pressure machines, etc. Some of them need prayer items like prayer mats, Tasbih, etc. Another fun fact is that a significant amount of the elderly are addicted to betel leaf and tobacco. So, from this bar chart, we can conclude that the most important items for the elderly are medicines, walking sticks, eyeglasses, newspapers, and betel leaves. We even found ideas regarding their medicine-taking habits. We found from Figure 1.6 that about 52.4% of our target elderly could not read the prescription properly. Some of them had Cataract syndrome, some were

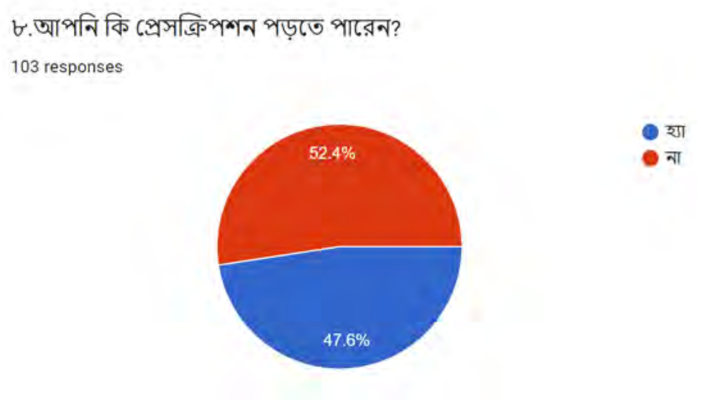


Figure 1.6: Pie chart of Bengali citizens prescription reading trend

illiterate and others had poor eyesight. Again, 55.3% of them declared that they could not remember where they kept their important items. About 34% of the elderly also said that they often tend to forget their respective prayer times. Figures 1.7 and 1.8 show some of our findings.

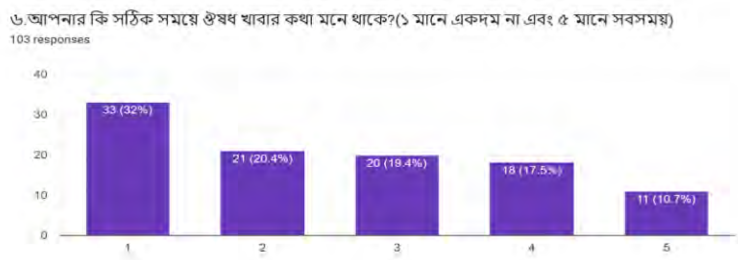


Figure 1.7: Our survey findings 03

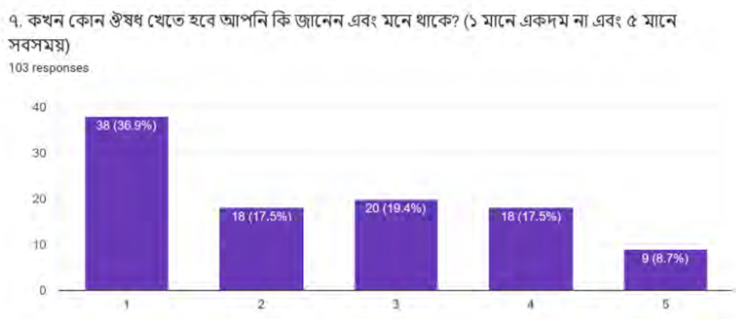


Figure 1.8: Our survey findings 04

1.3 Research Objective

The main objective of our research is to present modules for a system that will act as a friendly, simple, yet effective companion to the elderly. To achieve such a goal, we divided our objectives into three distinct parts.

1.3.1 Medicine Detection

Our medicine detection model could detect Bangladeshi medicines the user needs to take with utmost accuracy. We will initially train our model to detect most of the common medicines available for Bangladeshi elderly people on the market. Our model will detect the pre-trained medicines in real time using a webcam. Moreover, we are looking to detect if the user is taking the right medicine or not by the prescription that is stored for that specific user.

1.3.2 Reminder and alert the user

Another objective of one of our modules is to make the user more effective in his/her day-to-day life. We want to make sure to alert the user of certain works he/she needs to do timely such as; prayer, meeting, taking medicines, etc. We also want to remind him/her to do proper exercise. It will also be able to make crucial decisions like calling emergency contacts in case of such dangerous situations.

1.3.3 Human interaction and entertainment

As our targeted users live in isolation, we are looking to make our modules a user-friendly companion for them. Our system will interact with the user more like a friend. It will engage in human-like conversations with the user and keep him/her company. It will ask questions and give answers to the user's various questions. More importantly, it will reduce the boredom an elderly person feels in his/her day-to-day life. In different situations, our system will also provide recommendations and opinions to the user.

1.3.4 Elderly people fall detection and alert

Our targeted group is mostly elderly people in Bangladesh. They always live in constant fear of physical damage in everyday life. We want to monitor our elderly to let them live a risk-free life. We do not want the elderly to hurt themselves. So, we want to detect falls. To provide a comprehensive approach to aged care and management, future work will require advancing technology and working in conjunction with healthcare providers.

1.3.5 Emergency alert sender through voice recognition

An elderly person can have a severe condition and might not be able to get up from bed or unable to contact anyone when he needs help. This emergency alert sender will send an alert to their caregiver or guardian in this kind of scenario. The elderly person just has to ask for help verbally in Bangla or English and the alert will be sent. This module will recognize alarming keywords through speech recognition and send an alert while giving the user the benefit of changing or adding new keywords. This module will minimize the time that takes for help to arrive at the time of emergency.

Chapter 2

Literature review & Related Work

Robots serving as companions for elderly individuals bring quality benefits. These robots can significantly enhance medication management by detecting and reminding users about their medication schedules, ensuring timely and accurate intake. For those with complicated medication schedules or cognitive problems, this is beneficial. Monitoring their posture and identifying potentially dangerous positions helps to ensure their safety and well-being. The elderly people along with their caregivers can be more comfortable if they can get timely notifications. Also, detecting dangerous postures and prompt notifications can help prevent falls and injuries. These intelligent robots which are enabled with the ability to engage in conversations help to face social isolation and the challenges of loneliness. Robots with medicine detection, posture monitoring, and talking could greatly benefit the elderly. These aspects promote the overall well-being, security, and quality of life of elderly people. Recognizing medicine with color and text research focuses on the challenges faced by elderly people who have poor eyesight or have difficulty differentiating their medicines. The writers of the paper focused on the fact that pills and tablets may come in various sizes, shapes, touchings, and light conditions. All these conditions make it difficult to differentiate between similar types of medicines even for a normal functioning human being let alone the elderly. Visually affected people could not have access to color and text information, which makes the identification procedure more difficult. The paper proposes a methodology for pill recognition by focusing on color information extraction and text recognition. The dataset is divided based on the number of colors and texts imprinted on the pills. Pill color information is extracted through image segmentation, followed by statistical measurements like kurtosis and skewness. These measurements help determine the number of colors present on the pill surface. For text recognition, the paper describes the detection of expected text regions to ensure error-free text detection. The authors utilize the NLM RxIMAGE database, a high-quality image dataset, for developing, training, and testing the proposed system. The overall accuracy of the system for color determination is reported as 95.6%, while the text recognition accuracy is 81.32. Additionally, the review may discuss the significance of using quality datasets, such as the NLM RxIMAGE dataset, for the development and evaluation of image recognition systems [8] . Figure 2.1 shows sample image of NLM RxImage dataset. [2] Jinseok Woo and Naoyuki Kubota proposed a conversation system that enables verbal communication between a human and a robot partner using multi-modal perception in the paper "Conversation System Based on Com-

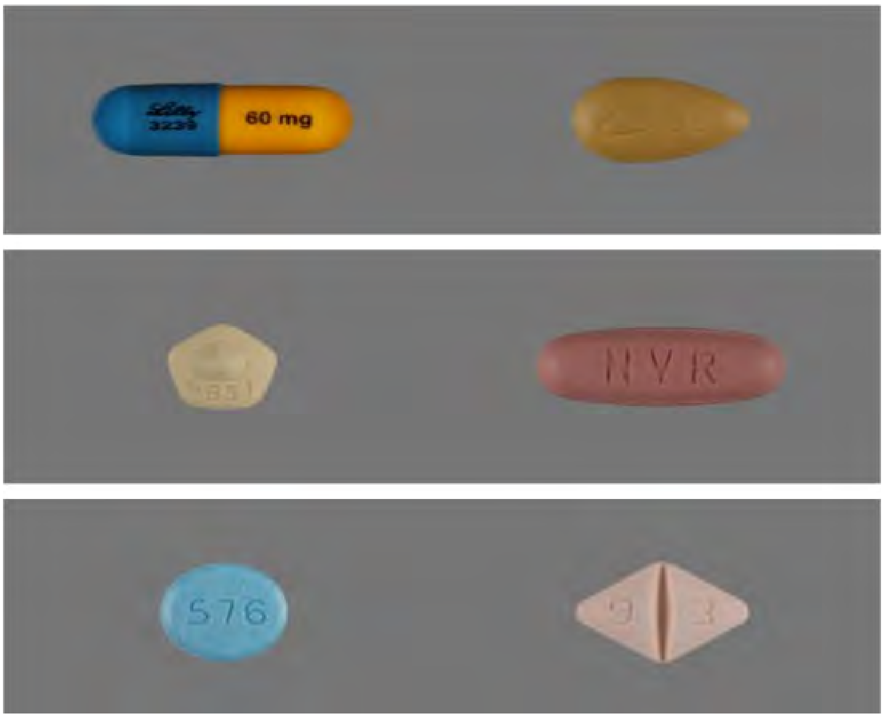


Figure 2.1: Sample image from NLM RxIMAGE dataset [8] .

putational Intelligence for Robot Partner Using Smartphone". The robot Figure is shown in Figure 2.2. The authors describe the control structure of the robot partner and its architecture, and they apply robot vision for human and object detection. Additional conversation system based on informationally structured space and a method of conversation learning based on human utterance patterns and perceptual information. The paper's findings show the proposed method's effectiveness in enabling natural communication between humans and robot partners. The authors show experimental results that show the robot's ability to recognize and respond to human utterances and perceptual information. The proposed conversation learning method allows the robot to incrementally learn and memorize utterance sentences, leading to improved conversation flow over time. However, the paper acknowledges some complications and limitations. It does not include syntax analysis, which limits the robot partner's understanding of differences in person and context. The authors suggest the need to incorporate syntax rules to enhance the reuse of memorized utterance sentences and the performance of the voice recognition system, and the automatic proofreading function has room for improvement. [7] Parvin, Paternò, and Chessa proposed an approach to detecting abnormal behavior by developing a profiling strategy using task models that specify normal behavior as anomaly detection in the daily behavior of the elderly has gained significant attention in recent years.The model architecture is shown in Figure 2.3. Their algorithm compares planned and actual behavior to identify deviations and categorize the anomalies. The resulting environment generates modal actions, such as reminders and notifications, based on the detected behavior to support the well-being of older people. Ambient intelligence technologies, such as sensors and interactive devices, are being used to support the everyday living of older individuals. These technologies enable the analysis of gathered data to identify people's current state and activities, providing a safe and secure environment for independent living. Monitoring

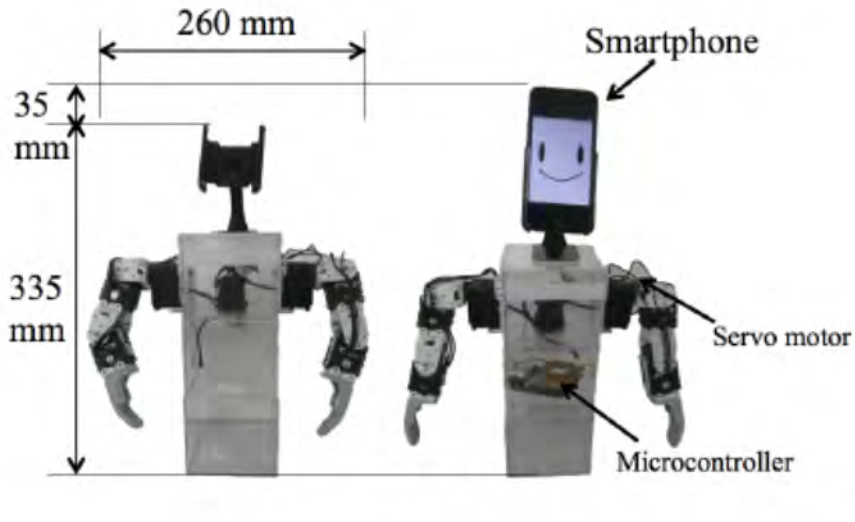


Figure 2.2: Robot partner : iPhonoid [2] .

daily activities, such as bathing, dressing, food preparation, and medication intake, is crucial for assessing an individual's ability to live independently. Task analytic methods have been employed in numerous studies to regulate human behavior. Enhanced Operator Function (EOFM), and ConcurTaskTrees (CTT) are examples of task analytic modeling systems. Task models offer a more compact representation of information, and transformation algorithms can generate Markov Models from task models for further analysis. Detecting anomalous behavior in real-time is challenging due to the streaming and time-series nature of the data collected from older people. However, these methods have limitations, such as the need for pre-classified data, complexity with large sets of transitions, slow training, or lack of a valid solution in certain cases. [4] Another study aims to enhance the robot's interaction capabilities, focusing on creating a more engaging and personalized experience for users. It seeks to identify the impact of this socializing ability on users' perception of the robot's intelligence, social acceptance, and likability. Neuro-Inspired Companion (NICO) robot, which utilizes face detection and tracking algorithms to engage users in personalized conversations. NICO recognizes individuals, learns their names and preferences, and remembers interactions. The interaction scenario, implemented using a state-based dialogue manager, allows users to communicate with NICO naturally and trigger learning scenarios where the robot can be taught new tasks. Preliminary results indicate that users perceive the socially interactive robot as more intelligent and likable compared to traditional button-driven robots. While the paper highlights the positive impact of enhanced interaction capabilities, several complications arise. The difficulty of developing humanoid robots that resemble people without causing the Uncanny Valley phenomenon is highlighted. The effectiveness and face detection and recognition algorithms in real-world scenarios requires further improvement. Furthermore, long-term user engagement and personalized interactions across a larger user base are a challenge for future research in Human-Robot Interaction.[1] Additionally, another project in the paper "A Posture Recognition-Based Fall Detection System for Monitoring an Elderly Person in a Smart Home" provides an overview of existing fall detection systems and high-

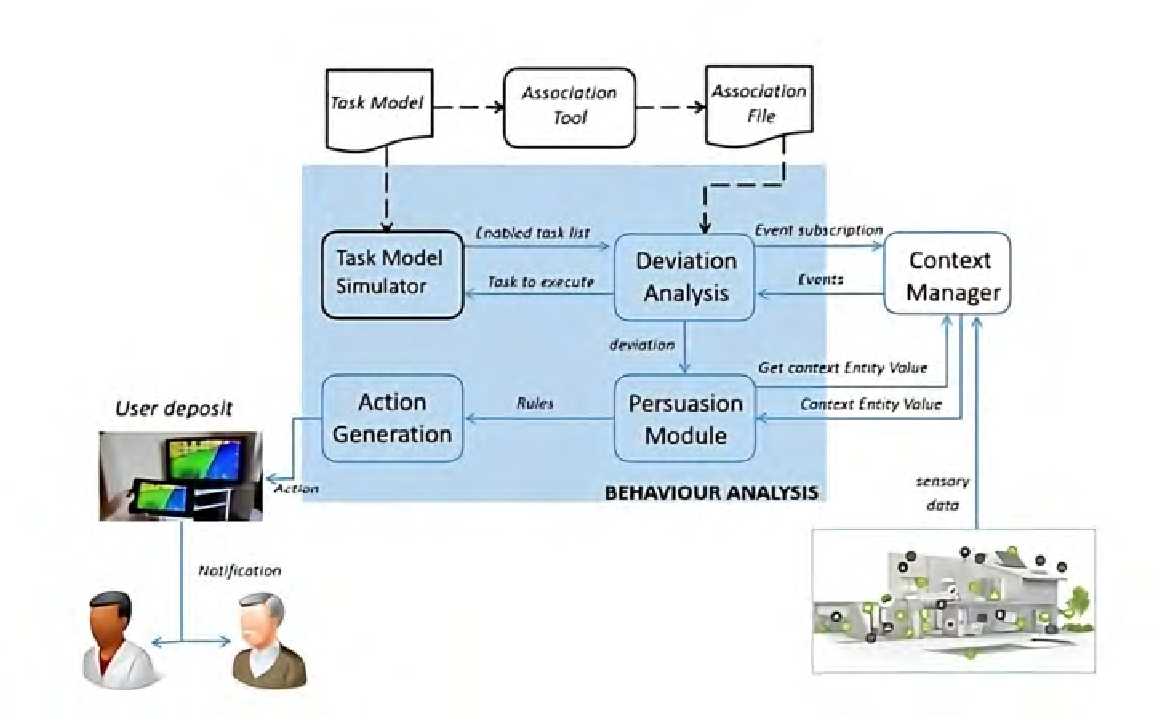


Figure 2.3: The architecture of the solution [7] .

lights the need for a computer vision-based approach to monitoring the elderly in a home care environment. Fall detection techniques include non-computer vision-based methods and computer vision-based methods. Non-computer vision-based methods use various sensors to capture movement and detect falls, such as acceleration sensors, acoustic sensors, and floor vibration sensors. While these methods have shown some success, they can be inconvenient for the elderly to wear and are susceptible to environmental noise, but Computer vision-based methods use computer vision and image processing techniques. These methods extract information from video recordings and apply algorithms to detect accidental falls. The proposed fall detection system in the research paper uses background subtraction to extract the human body region, followed by post-processing techniques to improve accuracy. Features extracted from the human body region, including ellipse fitting and projection histograms, are used for posture classification. A directed acyclic graph support vector machine (DAGSVM) is employed for classification, and floor information is integrated to enhance fall detection accuracy. Several challenges are faced when developing and implementing fall detection systems, particularly those based on computer vision like variability in human behavior, environmental factors, real-time processing, and privacy concerns. However, it acknowledges the need for further evaluation and emphasizes the system's applicability to monitoring a single person. [13] Furthermore, one more research paper titled "Artificial Intelligence-based Voice Assistant" explores the developing field of voice-controlled AI assistants, emphasizing their significant impact on daily life. The study primarily focuses on an AI-based Voice Assistant developed using the Python programming language. The assistant showcases a variety of features, ranging from playing songs and retrieving information from the internet to ensuring user safety, facilitating communication, and enhancing ease of life. The use of ASR demonstrates a methodical approach to voice recognition and clarifies the technological foundations. The strengths include the thorough description of the assistant's capabilities and the possibility for a wider range of applications, such as in house- hold appliances and education. The integration of ASR elucidates the technical underpinnings, showcasing a systematic approach to voice recognition and response. However, potential fragilities include the need for continuous improvement to enhance accuracy, especially in understanding complicated demands. Moreover, addressing privacy concerns regarding voice data and ensuring data security could further bolster the paper. Overall, this research contributes valuable insights into the practicality and versatility of AI-driven voice assistants, showing the way for im-proven user experiences and advancements in the field of AI-involved voice-controlled technology. Apart from this, in this paper [11] the creators describe the development of an Android smartphone application aimed to assist elderly individuals in providing easy living while ensuring timely and accurate medication intake. The main ideas of the paper revolve around using mobile technology to provide reminders and alerts for medication schedules, integrating voice-based notifications, and incorporating an emergency assistance feature. The system focuses on addressing the challenges faced by older adults in managing medication intake appropriately, aiming to reduce healthcare responsibilities and enhance the safety and security of elderly individuals. The methods utilized involve the development of an Android-based application incorporating features such as user registration, contact number entry, medicine intake entry, text-to-voice conversion, voice-based reminder alerts, and color-based service selection. These methods allow users to input their medication details and set reminders for timely doses, utilizing voice alerts for ease of use and accessibility. Strengths of this research include the focus on a pressing issue faced by the elderly population and the innovative use of mobile technology to address medication constancy. The application's user-friendly interface, voice-based reminders, and emergency assistance features enhance its usability for older adults. However, the paper could benefit from providing more details on the technological framework and implementation specifics. Additionally, an evaluation of the application's usability, effectiveness in medication adherence, and user satisfaction would add credibility and validation to the proposed system. Overall, the study shows a promising step that uses modern technology to support the elderly to manage their healthcare needs and improve their overall quality of life [10] We found another paper titled 'An Elderly Fall Detection System Using Depth Images'. We found a system using Microsoft's Kinect sensor. They acquired depth images through continuous video of Microsoft Kinect sensors that are mounted on a wall in a laboratory setting. The system utilizes continuous video for recording depth images. It contains background subtraction to single out human objects, ground segmentation for distinct fall poses, feature selection based on motion velocity, and quantitative assessment of ascertained falls using a decision tree algorithm. Besides the sitting, lying, and walking activities which constituted a cross-validation set used for testing of fall detection models sufficiently. [15] We also Vision-Based Elderly Fall Detection Algorithm for Mobile Robots (Raspberry Pie 4) They selected a simplified version of NanoDet called nanoDet-Lite to be used for fall detection as it is fast but low cost and deployments do not take much time. ShuffleNet V2 is $0.5\times$ complexity suited for real-time processing in this architecture. NanoDet-Lite is trained on the Python 3.7 and PyTorch environment that has been pre-installed in a server with an Nvidia Quadro M400 GPU, this therefore balances very well be-

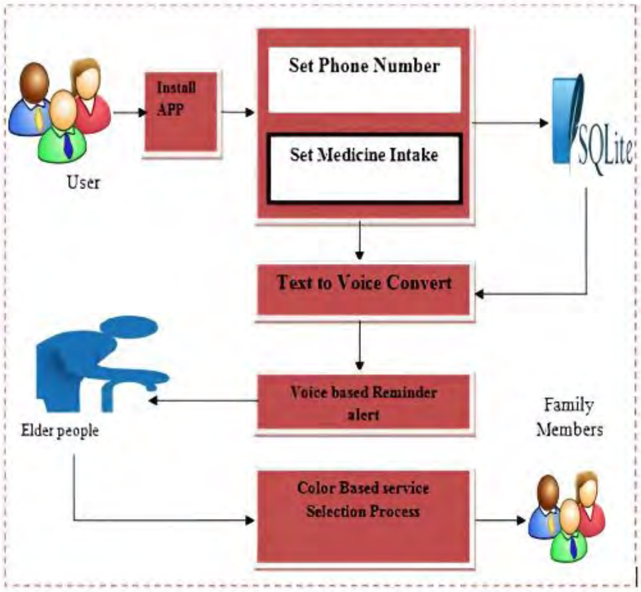


Figure 2.4: The Architecture Diagram of Proposed system [11] .

tween performance versus resource efficiency of computation. They used a dataset with 1275 images which combines both publicly available UR Fall Detection Dataset and their own. This variable dataset provides good training and validation. Limitations: The dataset is very limited in this case. [3] One more research paper titled 'An Integrated Vision-Based Approach for Efficient Human Fall Detection in a Home Environment' followed differently. Their proposed fall-detection and classification procedure includes five major steps:

- data preprocessing,
- image segmentation,
- feature extraction,
- fall identification using the GLR chart, and
- using the SVM algorithm for classification (see Figure 2.5).

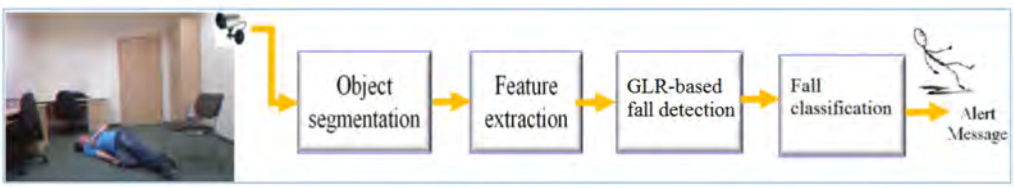


Figure 2.5: Diagrammatic representation of camera-based fall detection technique [3] .

Uses image segmentation to extract images from the camera. (Figure 2.6) In this work, they concentrated on non-zero pixels defining the human silhouette. five occupancy regions. These regions were acquired by a basic centring partition based

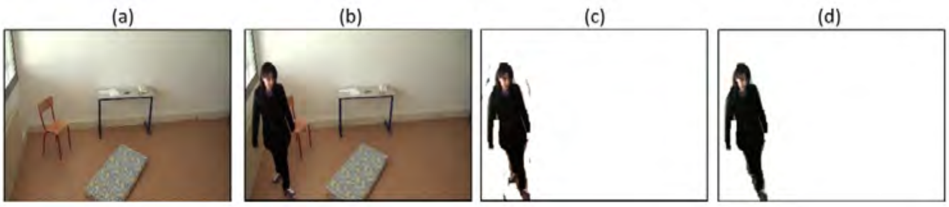


Figure 2.6: A sample of the image segmentation technique [3] .

Figure 2.7: Extracting features based on human body silhouette [3] .

on the body's centre of mass. The partitioning is done by setting five lines from the centre of mass of the body (Figure 2.7). Vertical on the first line, two other segments are placed at 45○ to either side of the vertical and further from them are applied two more tips located instead of relatively strong angles for the third and fourth. The five sections of Figure 2.7 3 d are coloured in cyan, green, yellow, blue and magenta respectively. Features are defined with the centre of mass and areas for each frame. Next, we compute normalized areas by dividing each subarea value, Ai; I D 1: 5. by the whole silhouette area:: The extracted features are adequate and pertinent. In addition, they vary regarding human postures and are cheap to calculate. The five ratios are calculated for each image and serve as input for fall detection and classification. (Figure 2.8). Figure 2.9 demonstrates typical

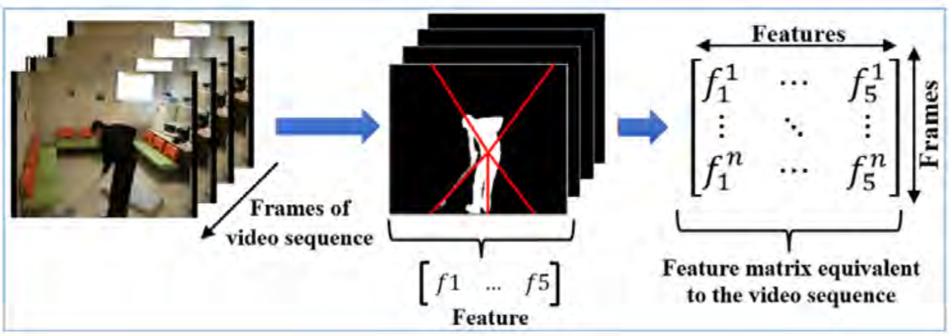


Figure 2.8: Feature used to detect a fall [3].

movements enacted by the inspected person while standing up, bending over a table to pick it up from the floor bottom, sitting on and lying down. Figure 2.9 reveals that Areas are more sensitive to fall events compared with other areas. It is observed by the size of space in these areas compared with plots found in other regions, demonstrating their suitability to be used when detecting falls. [16] Furthermore,

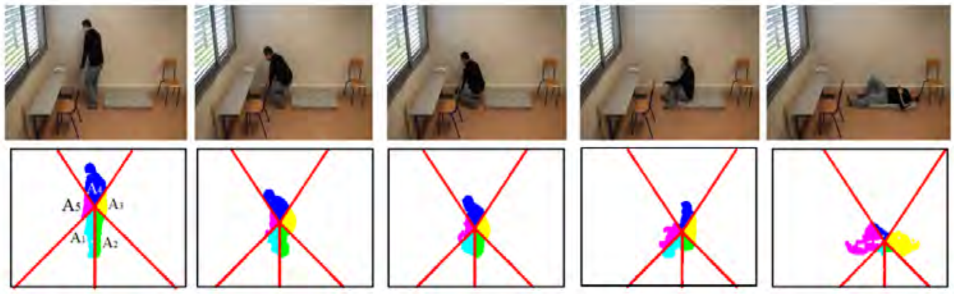


Figure 2.9: Samples of human body partitioning into five sections [3].

we found a paper named 'Fall Recognition Algorithm for the elderly based on home service robot'. This paper suggests an accidental fall recognition paradigm by the symmetry principle of home service robots. Here is a summary of the key points about their approach and limitations.

Approach:

- Detects video frames of the elderly person using a Kinect sensor on a home service robot.
- Applies OpenPose to obtain human joint keypoint information per frame.
- Uses four key parameters to identify a fall:
    1. Deceleration of the descent rate of the hip joint centre.
    2. The angle between the human body's centerline and ground plane.
    3. Ratio of the width and height associated with a bounding rectangle for humans.
    4. The fact that the person does not stand after falling.
- Confirmation of whether to put on an alarm can be done through voice interaction with the elderly person by the robot after a potential fall is detected.

Limitations/Drawbacks:

- Needs an in-house service robot with a Kinect sensor that may not be as popularly accessible or price-effective as at current.
- Only assessed on a limited set of 10 testers capturing the records of falls, which are not put into the context of larger reflections diverse real-world datasets.
- Uses threshold technique on the four central parameters to identify falling episodes. May not work for falls that happen in strange positions or trajectories which do really distinguish from expected patterns.
- No discussion on how occlusion is handled – falling detection could suffer if there are no key points because of an occluded object.
- Tested in a lab environment. Raluca Popescu The review asserts that how the elderly individual navigated given situations, performed activities of daily living such as bathing and cooking, sported glasses in office settings interacted with requests or addressed seeking questions remains unclear.

Putting it all together, the system proposed here to detect a fall uses the robot's onboard camera for a vision-based approach in building and detecting falls are summarized below. It is novel in its approach to home service robots but has limitations regarding fall patterns handled, robustness against occlusion and site testing within the real environment around homes as would be a more comprehensive evaluation on large real-world datasets. [14] Our final paper that we found was 'Human Fall Detection Algorithm Based on YOLOv3'. In this paper, the YOLOv3 algorithm is proposed, it has the advantages of high accuracy and low computational complexity.
Approach:

- Dropper lever uses the YOLOv3 object detection model to identify falls from video frames.
- K means clustering is used for optimizing the YOLOv3 anchor boxes, owing to its expansion on crucial features inherent in fall data.
- Builds an individual fall detection dataset of 2600 images.
- Trains a 3 model YOLOv3 from this dataset to recognize two categories – a person and a fall.

Results:

- Obtains 0.83M mean average precision (mAP) on the test set, including AP of 097 calls for the fall class.
- Faster R-CNN and SSD at PASCAL VOC databases outpaces other detection methods.

Limitations:

- About 2600 images were constructed in a relatively small database.
- Whereas he discusses uses for an exploratory model of fall patterns, he does not offer any analysis as to what falling pattern this model can and cannot be used.
- It does not address the performance in complex real-world conditions (occlusion, varied light etc.).
- Limited fall patterns observed during training may be too restrictive for the model to generalize from. The suggested solution is simply to consider whether or not this limited sample size, was available and necessary to ensure a good fit with Arthur Mendela's results; by using an unsupervised procedure called Hollywood Network mapping practitioners can create neural networks of entertainment ideas that exude creativity if they.
- They have no reported metrics regarding computation performance like frames per second.

Overall, the paper shows satisfactory accuracy with YOLOv3 for fall detection on a smaller custom dataset. There is no evaluation within larger real-world datasets, recognizing weaknesses of this approach and the need for its benchmarking with regards to computational performance that are necessary in cases when it would be applied during practical work. If these were addressed, their approach could appear more robust.

To conclude, the literature review has shown us that to build a system based on multiple modules for elderly people in Bangladesh, we can integrate several types of learning models to adjust to behavioural patterns and support their well-being. However, to fulfil their overall needs, we need to implement a complete package that can support their multitude of needs. This research targets a system that aids in supporting elderly people in medicine intake, physical awareness, and healthy interaction so they do not feel isolated.

Chapter 3

Work Plan

To keep our research focused on major issues that are faced by the elderly people of Bangladesh, we did surveys to figure out the most troublesome challenges an elderly individual would face. We planned to solve those problems by using the latest technology. For medicine detection, we collected prescriptions from elderly people to create a dataset of medicines they frequently took. We created the dataset of medicines and trained our model accordingly so that it would perform marvelously in real-time. For fall detection, we collected the dataset to train our model to detect any person falling immediately and send an alert to the caretaker. To mitigate the loneliness of an elderly individual, we made a voice bot capable of conversing in Bangla with that person. Also, the voice bot can be customized to talk about specific things. We did surveys to find out what kind of conversations an elderly person in this country would like to have and gave our voice bot the information to discuss it. The voice bot can be personalized according to user preferences. They are very forgetful people and might need assistance from their caretaker anytime. We created a reminder system to remind them of the important things from their schedule and an alert system that would send an alert to their caretaker. The alert system understands Bangla keywords through speech recognition for user benefit. Figure 3.1 shows our Overall Work Plan. Figure 3.2 shows the flowchart of the voice assistant. Figure 3.3 shows the flowchart of emergency assistant, Figure 3.4 shows the flowchart of fall detection and Figure 3.5 and Figure 3.6 shows medicine detection and reminder flowchat. Detailed workflow is given below.

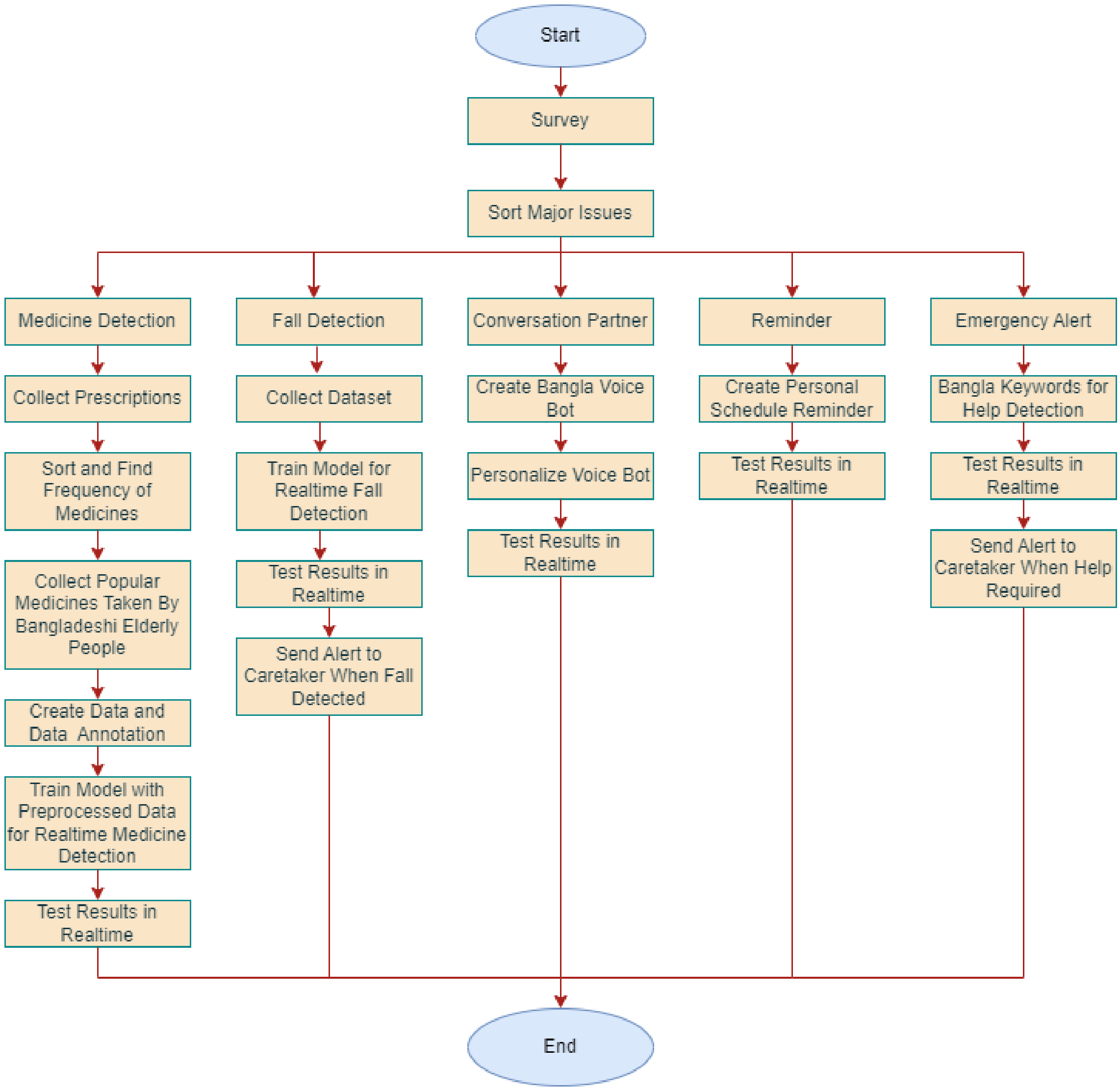


Figure 3.1: Overall Work Plan.

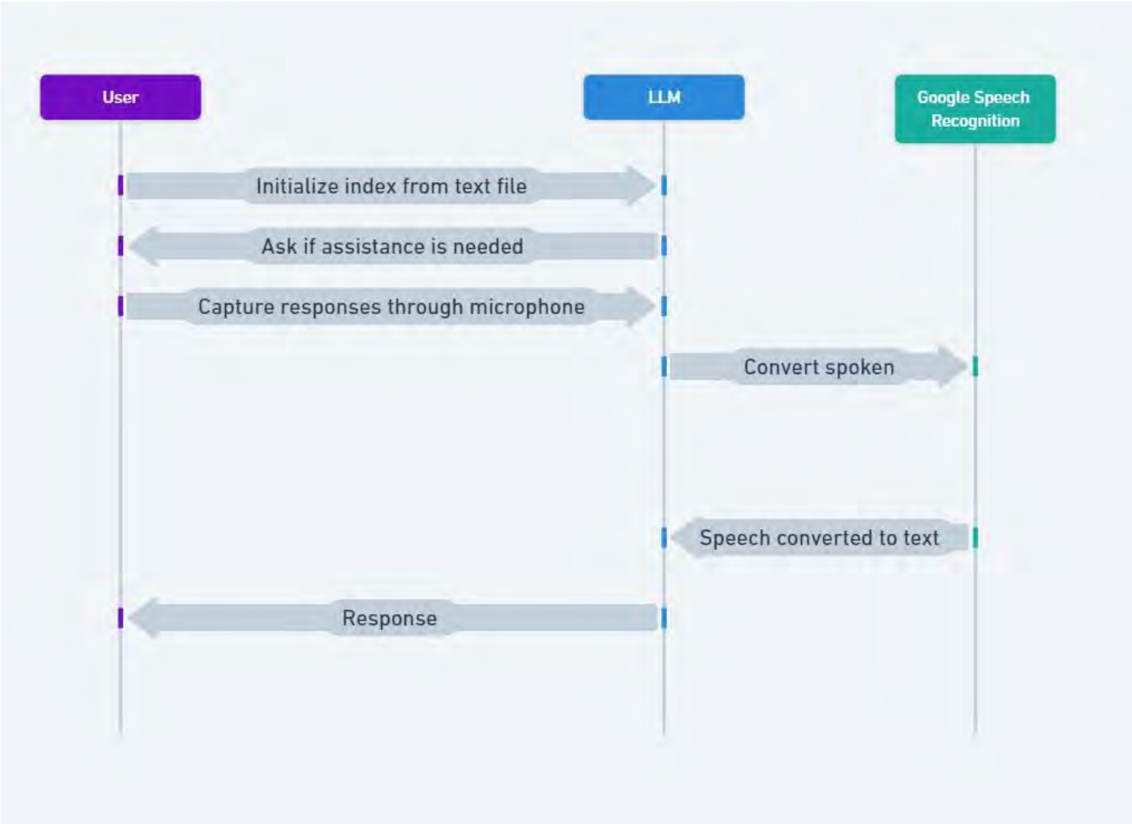


Figure 3.2: Flowchart of voice assistance.

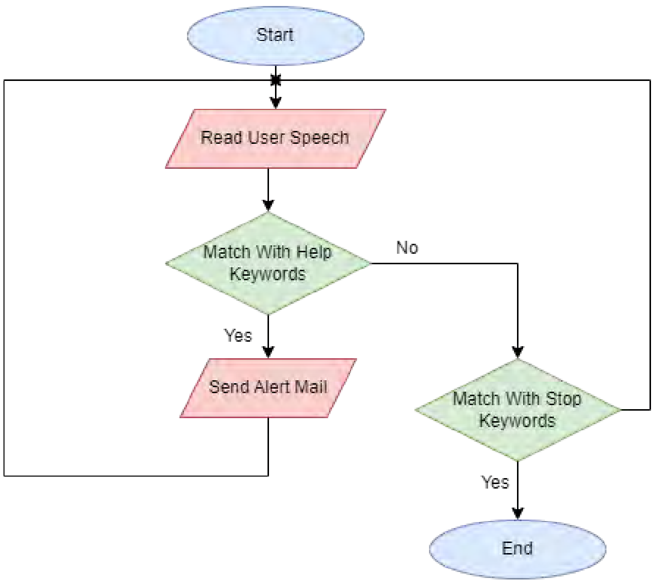


Figure 3.3: Flowchart of Emergency Assistance.

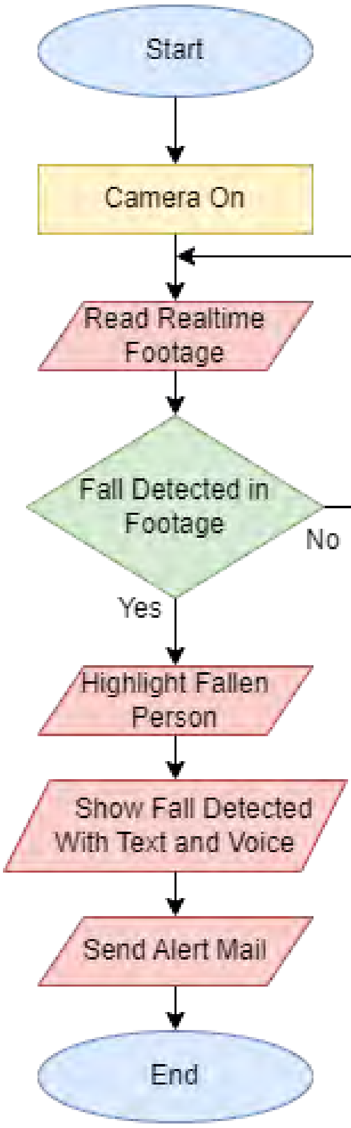


Figure 3.4: Flowchart of fall detection.

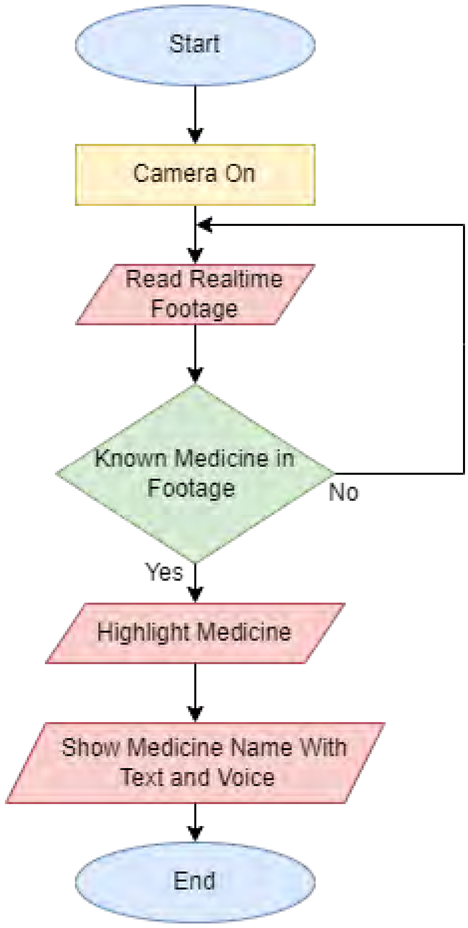


Figure 3.5: Flowchart of medicine detection.

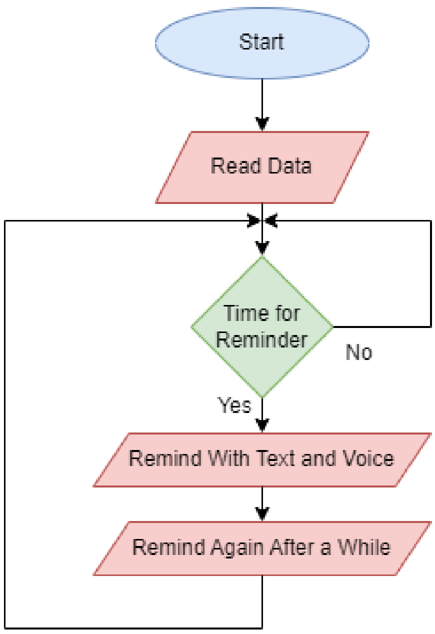


Figure 3.6: Flowchart of reminder.

Chapter 4

Detailed Description of The Prototype

4.1 Medicine Detection

In this process, when a user seeks information regarding a specific medicine, the system seamlessly integrates the YOLOv8 algorithm. This cutting-edge algorithm is deployed to meticulously inspect whether the queried medicine exists within its pre-existing dataset of recognized medicinal entities. The advanced capabilities of YOLOv8 allows for swift and precise identification of the medicine in question. Upon successful identification, the system proceeds to furnish the user with a comprehensive summary of details about the specified medicine. This information is conveyed through a dual modality of communication involving both articulate voice output and detailed text-based information.

4.1.1 Dataset

The importance of data in improving AI models is clear, especially with the rising popularity of machine learning. When creating a smart medicine detection system, we faced some challenges in getting the right data. Our main set of 2961 images was all put together by our team. We carefully took and worked on each picture to make a special collection that's the foundation for training our system. This set focuses specifically on the details of medicine and the details found on the backside of medicine strips. It's important to highlight that we didn't use any outside help; every picture in our set comes from our work. Additionally, it's worth noting that we took images from 21 different medicines to create this dataset, showcasing our commitment to diversity and comprehensiveness in our research efforts. Figure 4.1 shows our raw dataset.

Figure 4.1: Medicine Pictures for Data annotation.

4.1.2 Data Collection Process

In the systematic process of data collection for our medicine detection algorithm, we began by directly obtaining prescriptions from individuals aged over 40, amassing a total of approximately 200 prescriptions. With a focused approach, we selected prescriptions from around 165 individuals to ensure a representative dataset. Figure 4.2 shows some of our collected prescriptions. Personal details have been blurred. To enhance organization, we sorted the prescriptions based on the generic names of the prescribed medicines, enabling a more systematic categorization. Table 4.1 shows the data collection from the prescriptions. Patient names have been blurred for privacy. Employing a frequency-based strategy, we identified the most commonly prescribed medicines within the dataset. (Shown in Table 4.2) The final step involved selecting the most frequent 21 medicines, ensuring a diverse and comprehensive dataset for training our algorithm. This methodical approach demonstrates our dedication to capturing a purposeful dataset that aligns precisely with the objectives of our research in medicine detection for elderly people.

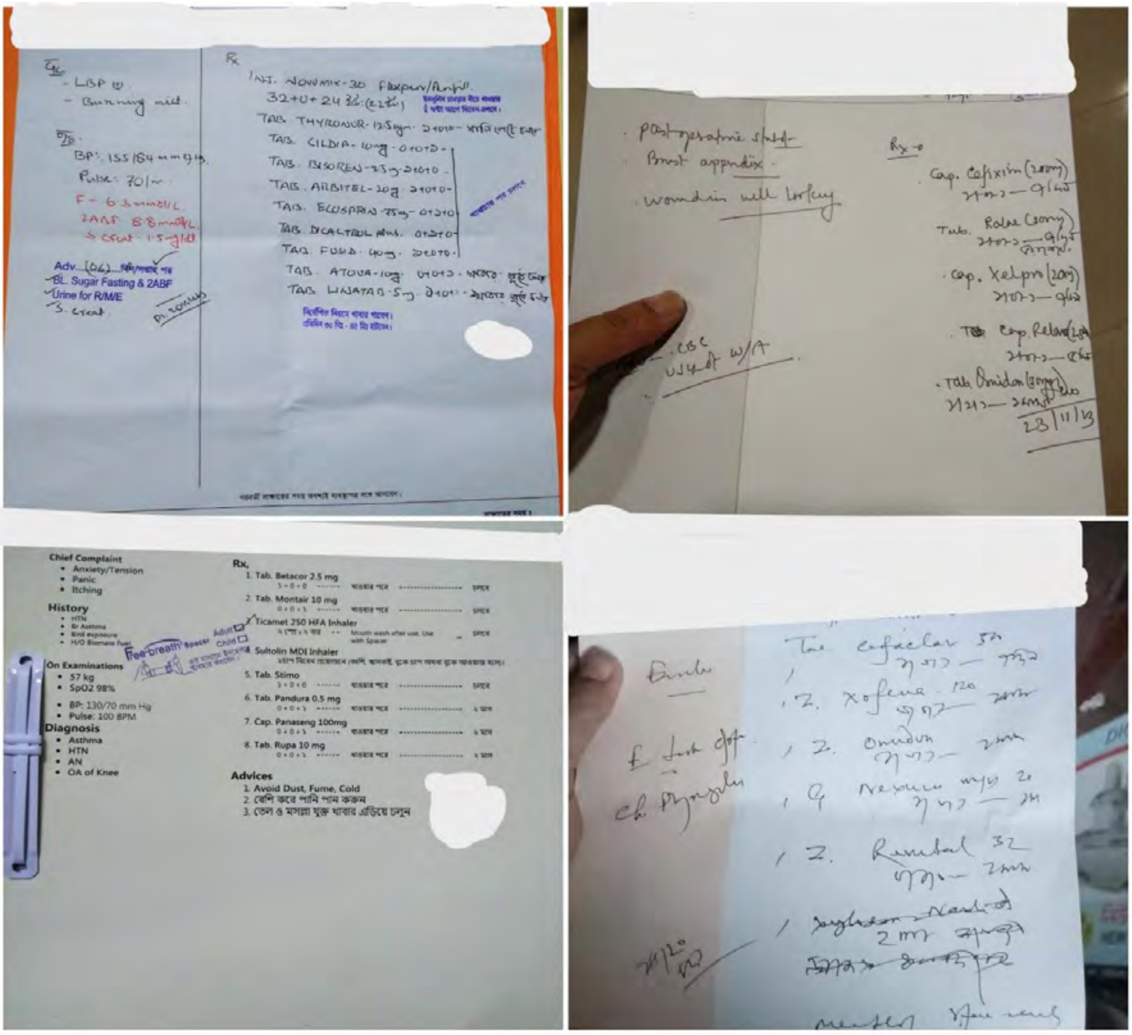


Figure 4.2: Collected Prescriptions.

Name	Age	Medicine Name	Generic Name
	65	Thyronor 12.5	Levothyroxine Sodium
		Cildip 10	Cilnidpine
		Bisoren 2.5	Bisoprolol
		Arbitel 20	Telmisartan
		Ecosprin 2.5	Aspirin
		Dicaltrol plus	Calcitriol + Calcium
		Fosid 40	Calcium Carbonate + Vitamin D3
		Atova 10	Atorvastatin
		Linatab 5	Linagliptin
	56	Betacor 2.5	Bisoprolol
		Montair 10	Montelukast
		Stimo	Flupenthixol + Melitracen
		Pandura	Clonazepam
		Rupa 10	Rupatadine
	70	Beklo	Baclofen
		Dicaltrol plus	Calcitriol + Calcium
		Bizoran 10 + 40	Amlodipine 10 + Olmesartan Medoxomil 40
		Gaba	Gabapentin
		Neuro B	Vitamin B1 + Vitamin B12 + Vitamin B6

Table 4.1: Organizing Prescription Data

Generic Name	Frequency
calcium + vitamin d3	49
esomeprazole	44
clonazepam	27
domperidone	25
iron + folic acid + zinc	25
multivitamin	23
dexlansoprazole	22
cefuroxime	21
vitamin b1 + vitamin b12 + vitamin b6	18
baclofen	18
ketorolac tromethamine	16
paracetamol	15
aspirin	15
flupenthixol	14
bisoprolol	13
montelukast	13
amlodipine	12
vonoprazan	12
pregabalin	11
pantoprazole	11

Table 4.2: Medicine Selection based on Frequency

4.1.3 Difficulty in Gathering Information

The process of constructing our dataset for medicine detection was not without its share of challenges. One significant problem we faced was collecting the prescriptions directly from individuals aged over 40. This task required asking for consent, coordinating with healthcare providers and following ethical rules of research. The number of prescriptions, totaling around 200, was a complex organizational challenge, conducting an organized approach to arranging and managing the data. Additionally, we needed to select prescriptions from approximately 165 individuals to gather medicine names of different varieties, involving communication, time management and data storage. One notable challenge we faced during the data collection process was understanding the hand written prescriptions, as some prescriptions became unusable. This took legibility from some of the prescriptions and affected the quantity of the information we gathered for our dataset. Even though we had to tackle these challenges, our commitment to produce a dataset of variety and usability remained fixed.

4.1.4 Data annotation

The data annotation process was one of the most critical parts of refining the dataset for medicine detection. We used Roboflow as our annotation tool. We had 2961 images that we needed to annotate. We used polygon bounding boxes instead of traditional rectangular ones. We used the backside of the medicine strip which helped to detect the correct medicine precisely. The polygon boxes helped us to annotate when the medicine strips were irregularly shaped accurately, minimizing the space within the bounding box, potentially improving IoU calculations. Furthermore, the polygon bounding box helped to deal with the scenarios where the tablets were partially visible. It can also adapt to objects in different orientations. Figure 4.4 shows polygon bounded box through Roboflow. With precise annotation and by the use of polygon bounding boxes our medicine detection dataset was prepared to be utilized properly for training.

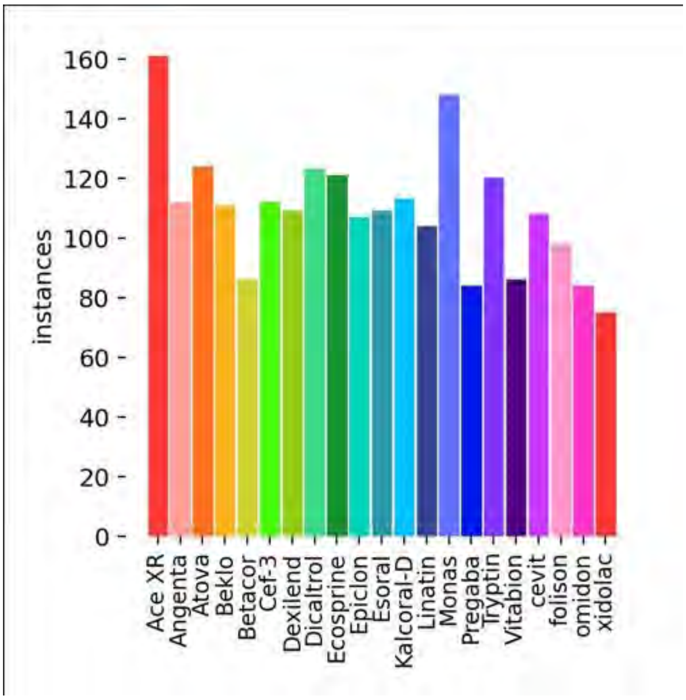


Figure 4.3: Labeled Classes.

4.1.5 Preprocessing

Preprocessing is very important to optimize data for machine learning models. It makes sure that input data is suited for training and gives accurate predictions. In our research, prepossessing is very important to enhance the quality of our dataset. It includes resizing the images uniformly with the standard dimension of 640*640 pixels. It provides a consistent format, facilitating an effective training model. It reduces the computational complexity and potentially increases the model training time. The advantages of these preprocessing steps are so many. It increases computational efficiency and helps in consistent data representation. This helps to train a more effective and stable medicine detection model.

4.1.6 Dataset Split

Another Important stage of data preparation is dataset splitting. It helps to evaluate performance, reduce overfitting and generalize the data. We were careful while splitting our dataset. We allocated 77% of our images totaling 2277 images for training. This provided a good number of images to teach our model the intricacies of medication detection.21% of our dataset having 618 images was used for validation.

Figure 4.4: Polygon Bounding Box Through Roboflow.

This helped us to fine-tune our model during training and optimize performance. Lastly, we used 2% of the dataset having 66 images for testing. Shown in Figure 4.5. We kept a lower percentage of our data for testing because we planned to test it extensively in real-time. This helped us to make sure that our model is robust and adaptable to real-time.

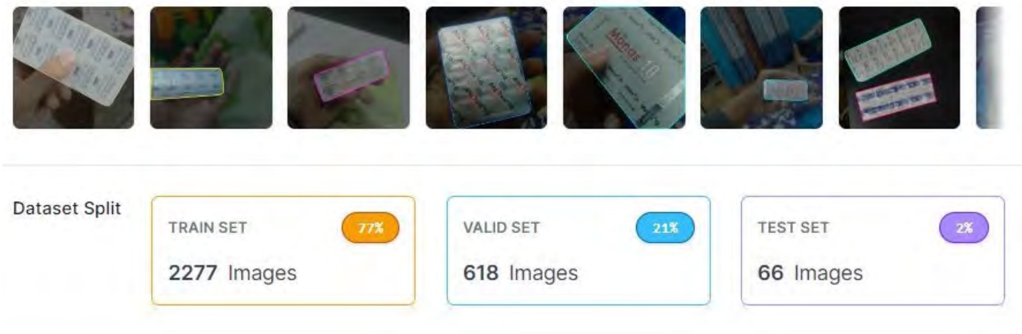


Figure 4.5: Dataset Splitting.

4.1.7 Model Description (YOLOv8)

The YOLOv8 model stands as an extraordinary innovation within the works of computer vision. It is a transformative evolution that builds upon the significant achievements of its YOLO predecessors. Differentiated as the latest iteration, YOLOv8 seamlessly connects innovative features and enhancements. As a result, it pushes the boundaries of both performance and adaptability to new higher stages. Its cutting-edge performance stands for a significant advancement in the field, assuring its status as a state-of-the-art solution with significant opportunities for the

future of computer vision research and application. The model is built to stand for speed, precision and comfortable utilization, thereby positioning it as a leading choice across the dimensions of complicated tasks. These tasks are not only limited to object detection and tracking but also include instance segmentation, image classification, and pose estimation. The collection of these qualities makes YOLOv8 and must be used in modern computer vision applications, promising to solve real-life scenario problems significantly.
Improved Speed: YOLOv8 differentiates itself due to its improved speed compared to its previous version which is YOLOv7. Speed is an important feature for practical real-time object detection and YOLOv8 performs very well in this aspect, providing accurate detection without having to affect the accuracy of the model. The processing speed takes a significant jump in efficient object detection algorithms, making it particularly a good choice for applications where fast identification is necessary.
Augmented Precision, Especially for Small Objects: YOLOv8 shows very improved accuracy, especially in detecting small-scale objects—an area where its previous version, YOLOv7, falls short when we compare them. This step forward is credited to the innovative implementation of a layered head network within YOLOv8. This effectively improves the accuracy of object detection. By focusing on improved precision for smaller objects, YOLOv8 uplifts the overall success and reliability of object detection so that it can serve a wider range of use cases.
Innovative Anchor-Free Architecture: A notable difference compared to its previous versions, YOLOv8 uses an anchor-free architectural paradigm. This change from traditional anchor boxes streamlines the training process, simplifying the model's adaptability to various types of datasets. The elimination of anchor boxes marks a significant paradigm change and it contributes to making the model flexible and simple to implement.
Strategic Multi-scale Prediction: YOLOv8 strategically uses a multi-scale prediction approach. It enhances the precision and versatility of object detection. By extending its predictive implementation across various scales, the model provides usable performance at identifying objects of different sizes, shapes and colors with improved accuracy. This calculated increase of quality in prediction methods highlights YOLOv8's commitment to achieving viable performance so that it can detect different types of object sizes and types.
Empowered by an Advanced Back- bone Network: YOLOv8 uses a more efficient and updated backbone network. It results in more improvement of the fundamental architecture design. This advanced network structure significantly impacts the model's object detection. It is even viable in challenging and complex environments. Updating the backbone network of YOLOv8 reinforces its ability to accurately identify objects, even in complex visual conditions. Because of this, the algorithm could take a significant jump forward in object detection functionalities.

4.1.8 YOLOv8 Network Architecture and design

The architectural design of YOLOv8 represents a significant update from previous versions in the YOLO algorithm series. Its foundation lies in a convolutional neural network. This is divided into two important components: the backbone and the head. YOLOv8 uses an enhanced CSPDarknet53 architecture as its backbone which is updated from YOLOv7 [20] features 53 convolutional layers and employs cross-stage partial connections to make the information flow between layers much faster. The head of YOLOv8 includes a series of convolutional layers followed by fully connected layers. These layers are responsible for predicting crucial parameters such as bounding boxes, objectness scores and class probabilities for detected objects within an image. Moreover, YOLOv8 introduces a self-attention mechanism in its head and it allows dynamic focus on various parts of the image and adjustment of feature importance based on tasks given. Furthermore, the model performs well in multi-scaled object detection and utilizes a feature pyramid network to correctly detect objects of different sizes and scales. This enables the processing of an image through multiple layers designed for this purpose. This collection features YOLOv8's ability to detect objects efficiently. Figure 4.6 shows the model architecture of YOLOv8 model.

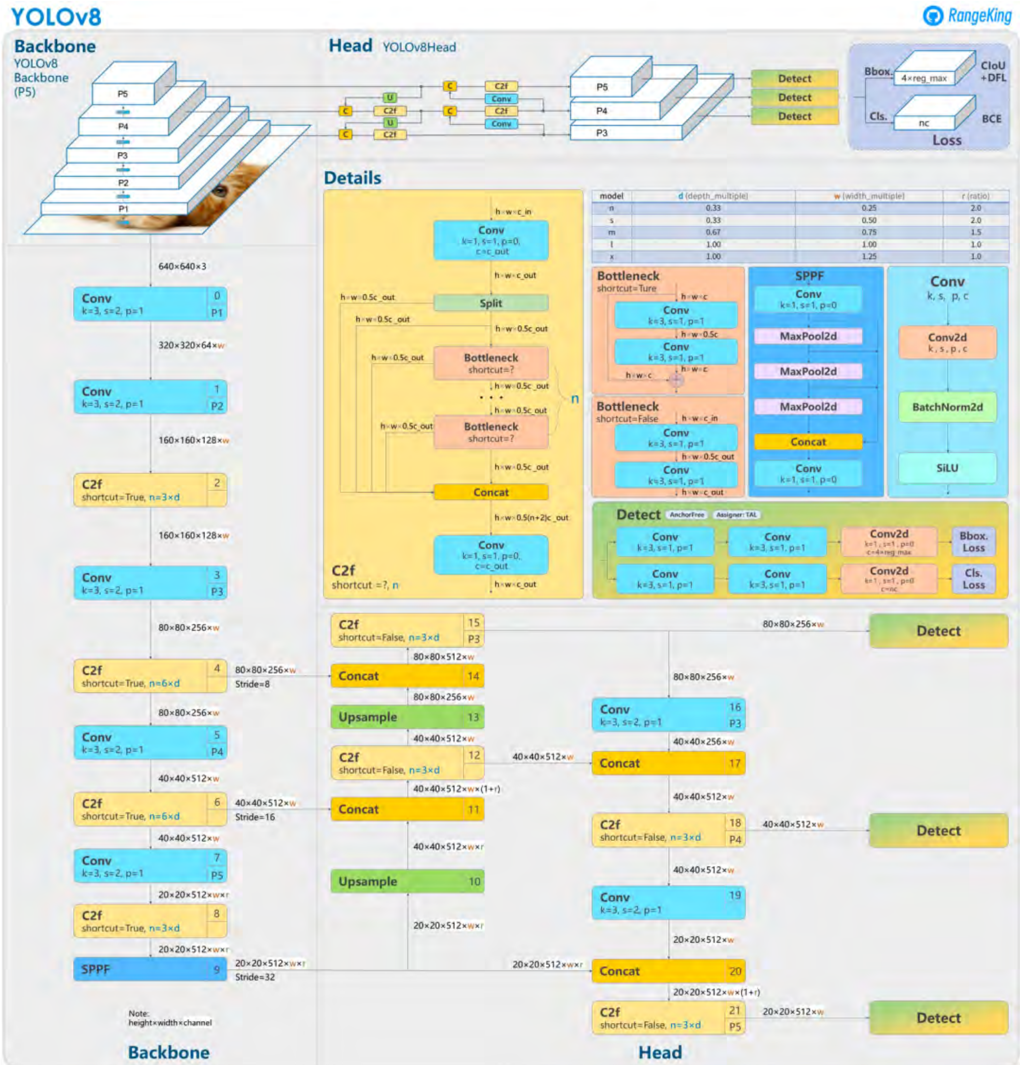


Figure 4.6: Model Architecture of YOLOv8 [21] .

To make YOLOv8 capable of object detection the process of training is implemented utilizing a labelled image dataset. This process is a must-step for supervised learning. It allows the model to recognize and localize objects within images. The training dataset is prepared with images annotated with bounding boxes. The bounding boxes include outlining object locations and associated class labels. Through following these steps we provide the essential guidance for the model's learning process. Data Preprocessing: First we prepare the data which was annotated and resized for training. It was resized with the model's input dimension and standard color. Various data augmentation techniques were used to augment data.

Loss Calculation: The YOLOv8 model computes a composite loss function in the training phase. It is vital for effectively training the neural network.

Confidence Loss: This indicates the model's confidence in detecting if the object is present within the predicted bounding box. The training process of YOLOv8 has the knowledge and information to make precise predictions on unseen data. This is a Robust and reliable tool for real-world object detection tasks.

Post-Training Object Detection Process for YOLOv8: For real-time detection of the strip within our dataset the YOLOv8 cross check with the trained image. This allows us to detect objects in real-time.

Forward Pass and Feature Extraction: Firstly it passes the unseen image through the trained model network. During forward pass this image traverses the network layer. This allows the model to extract complex features and relevant information .

Prediction Generation: YOLOv8 generates predictions for the input image. YOLOv8 predicts the bounding box of the medicine strips which is potentially within the strips.

Objectness Scores: Each polygon bounding box has an objectness score. That ensures the probability of an actual medicine strip being present within the box.

Class Probabilities: The model calculates class probabilities for each bounding box, indicating the likelihood of the object belonging to predefined classes representing different types of medicines.

Post-processing for Refinement: Following prediction generation, a crucial post-processing phase refines the raw predictions.

Confidence Thresholding: Predictions with low objectness scores are filtered out using a confidence threshold, retaining only the most confident predictions. This significantly improves the precision of medicine strip detection.

Non-maximum Suppression (NMS): Overlapping bounding boxes are addressed through NMS, which selects the most relevant bounding box for each detected medicine strip. This step minimizes redundancy and enhances the accuracy of object localization.

Bounding Box Refinement: Further refinement of bounding box coordinates is performed to align them more accurately with the true boundaries of the detected medicine strips.

Final Object Detection and Annotation: The refined predictions yield the final detected medicine strips within the image. These detected medicine strips are annotated with accurate bounding boxes specifying their locations and corresponding class labels denoting the type of medicine they represent.

4.1.9 Comparison with Other YOLO Model

The comparative analysis in the graphical representation encompasses four specific models: YOLOv8, YOLOv7, YOLOv6-2.0 and YOLOv5-7.0. Notably, YOLOv8 emerges as the swiftest performer but concurrently exhibits the lowest Mean Average Precision (MAP). Conversely, YOLOv5-7.0, while being the slowest in terms of speed, boasts the highest MAP among the models under scrutiny. The comparative analysis in the graphical representation encompasses four specific models: YOLOv8, YOLOv7, YOLOv6-2.0 and YOLOv5-7.0. Notably, YOLOv8 emerges as the swiftest performer but concurrently exhibits the lowest Mean Average Precision (MAP). Conversely, YOLOv5-7.0, while being the slowest in terms of speed, boasts the highest MAP among the models under scrutiny.

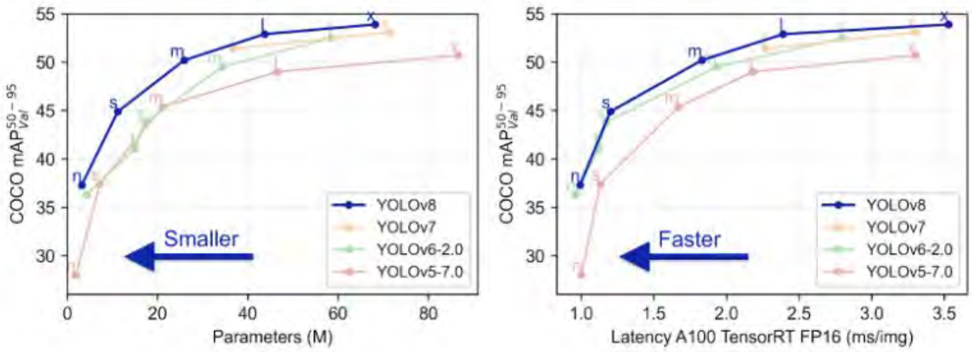


Figure 4.7: Comparison between YOLO models [21] .

The decision on which model to employ hinges upon the precise requirements of the application at hand. When expeditious object detection holds paramount importance, opting for a faster model becomes a viable choice. However, if pinpoint accuracy in object detection remains the primary objective, even at the expense of processing speed, then a slower but more precise model emerges as the optimal selection. This comparison underscores the critical trade-off between speed and precision and helps to choose the most suitable model.

4.1.10 YOLOv7

In comparing different models, we first used YOLOv7 to train for medicine detection. We went through 100 training sessions to help the model learn and get better at spotting medicines in images. The images were of the backside of medicine strips. YOLOv7 is known for being good at real-time object detection, making it a solid choice for our research. The 100 training epochs allowed the model to learn from our dataset. This improved its ability to accurately identify and categorize medicines based on the images we provided. This trained model is a key point of comparison. This helped us understand how well YOLOv7 works for medicine detection. Figure 4.8 shows the YOLOv7 training result and Figure 4.9 shows the normalized confusion matrix for YOLOv7.
Train/box loss and validation/box loss both of these curves decrease steadily throughout training. This indicates that the model is learning to predict bounding boxes for backside strips with good accuracy. The gap between the curves remains small, suggesting that the model is not overfitting to the training data. Train/class loss and validation/class loss are similar to the box loss, both classification loss curves decrease over time. This shows improvement in identifying backside strips. The slight divergence between the curves in the later stages could be worthwhile for further investigation and experimentation. Train/dfl loss curve is specific to YOLOv7 with a Dilated Encoder head. This also shows a steady decrease, suggesting effective refinement of bounding boxes. The model shows good progress in learning to detect and classify backside strips. It includes both loss curves and performance metrics

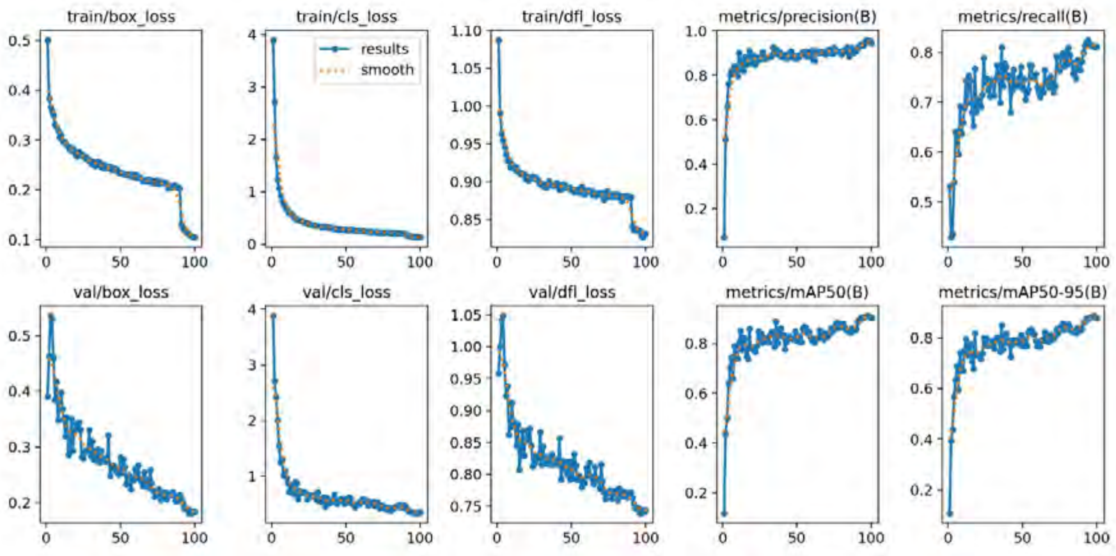


Figure 4.8: YOLOv7 Training results.

with positive results. The precision is high, indicating that most predicted strips are true positives but the recall could potentially be improved. This means the model might miss detecting some actual backside strips in the images. However, it should be mentioned that when multiple medicines are shown in a single image, the model experiences challenges. Those challenges include a decrease in accuracy and occasional failures in detection. This failure of stability underscores the complexity of the task. So we can reach a point where it is notable that further refinement is needed to enhance the model's ability to accurately identify and classify backside strips, specifically when we want to detect more than one medicine.
In the evaluation of our object detection model, we observed promising performance metrics, demonstrating the model's efficacy in accurately identifying and localizing objects. The precision (B) metric, representing the ratio of correctly predicted positive instances to the total predicted positives, reached an impressive value of 0.90. This highlights the model's ability to minimize false positives, which is crucial for applications where precision is paramount. The recall (B) metric, indicating the proportion of true positives captured by the model, achieved a commendable value of 0.80, showcasing the model's capability to identify the most relevant instances effectively. Furthermore, the mAP@50 (B) and mAP@50-95 (B) metrics which stand for the average precision over different intersection-over-union thresholds, resulted in viable overall performance with values of 0.88 and 0.84.

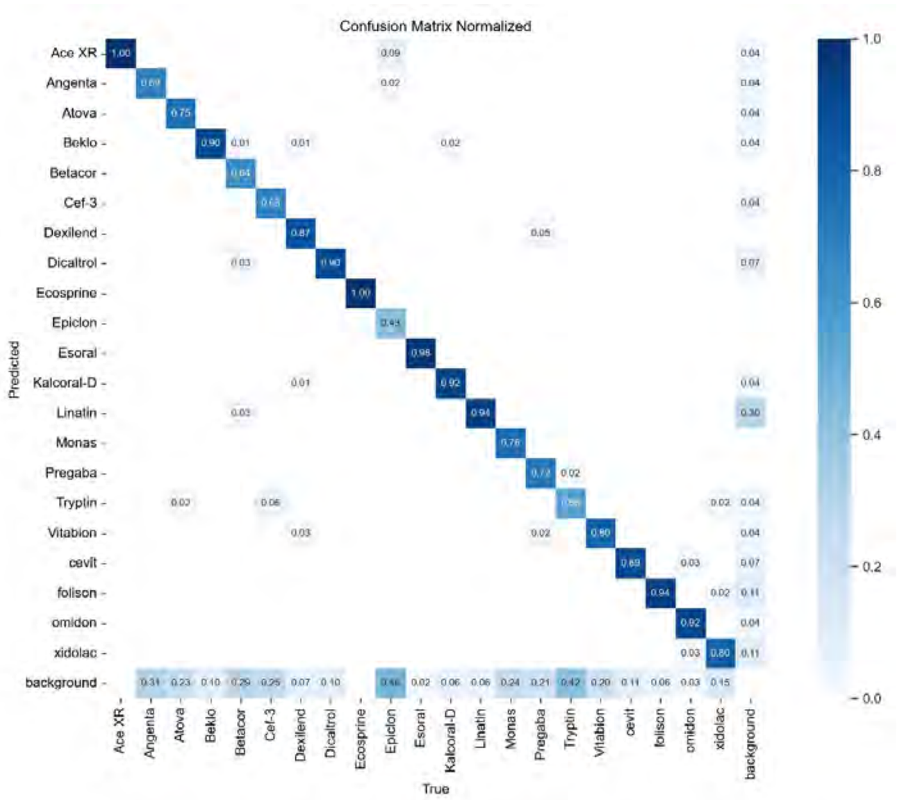


Figure 4.9: Normalized Confusion Matrix for YOLOv7.

Figure 4.11: Labeled Image (YOLOv7).

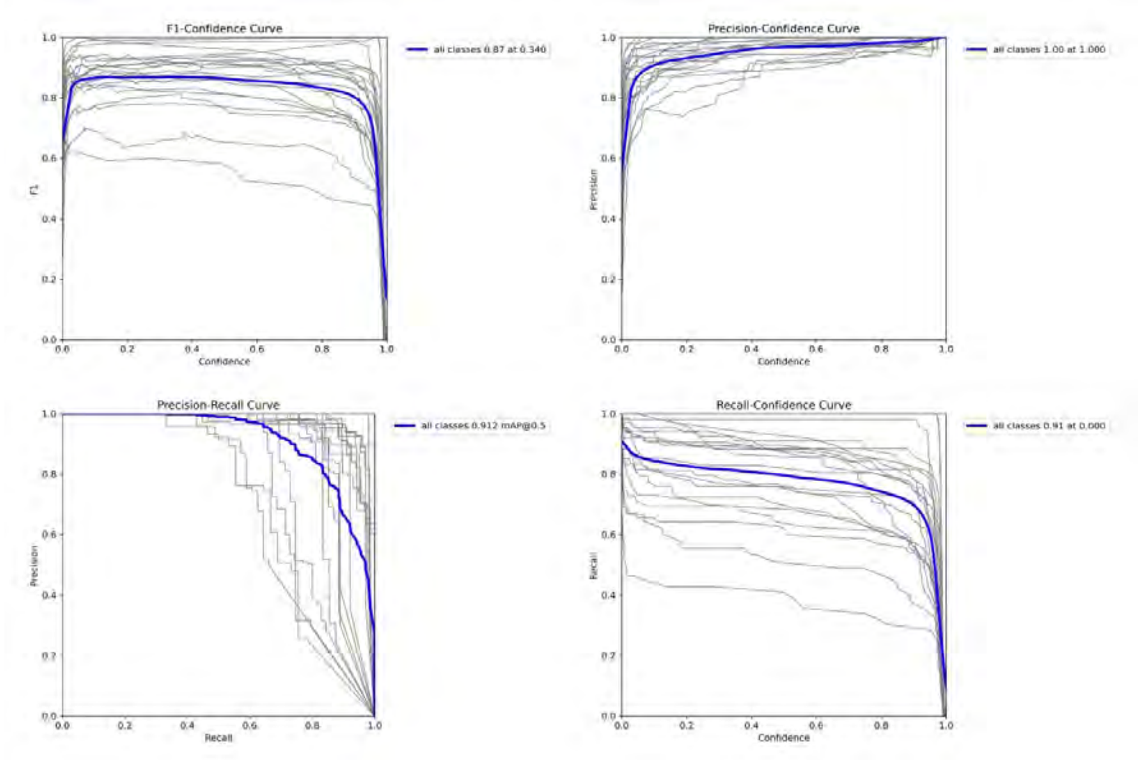


Figure 4.10: Performance Metrics for YOLOv7.

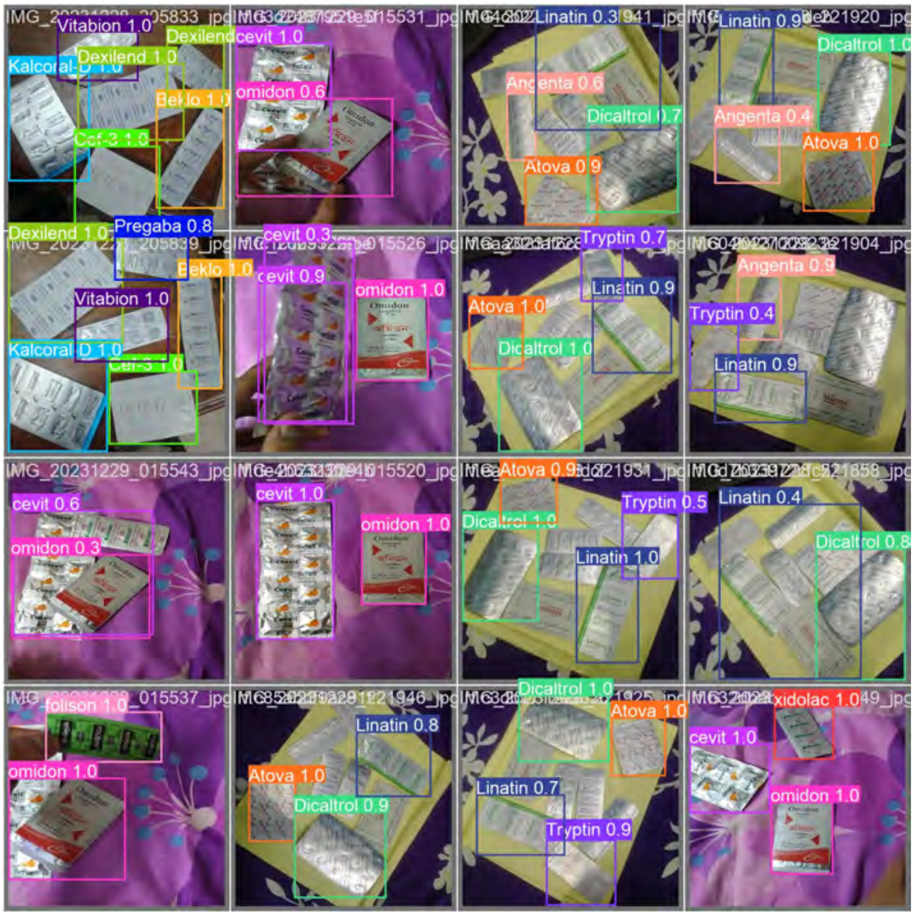


Figure 4.12: Predicted Image (YOLOv7).

4.1.11 YOLOv8

The YOLOv8 model is a better version of the popularly known and already famous for its high accuracy in object detection and speediness in terms of object detection. The YOLOv8 model has many improvements like better feature extraction techniques, improved optimization algorithms and enhanced training methodologies. 30 training epochs in our dataset enabled our model to learn and now, the YOLOv8 object detection model has reached greater levels of accuracy within a shorter amount of time compared to its predecessor the YOLOv7. The YOLOv8 model is one of the breakthroughs because it has shown great performance in real-time object detection and classification.

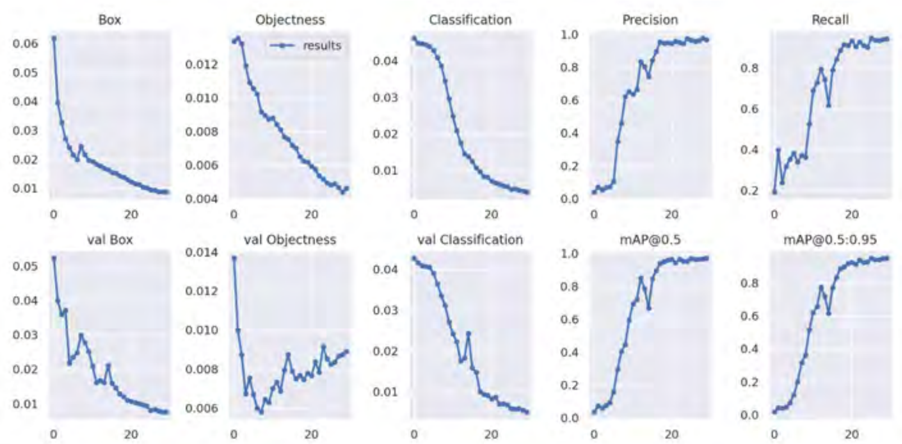


Figure 4.13: YOLOv8 training results.

Objectness and Classification Losses: The downward patterns in these losses suggest that the model is increasingly becoming more precise in predicting the bounding boxes, discriminating between background and foreground(objectness), as well as categorizing objects. This indicates that the model is learning to be more precise and confident in predicting locations as well as categories of objects present within an image.
Precision and Recall: The upward trend of the model's learning is rising trends in precision and recall metrics. Increasing accuracy means that the model is making fewer false positive predictions, while high recall implies an increased number of true positive cases captured. Balance is very important to have a strong and dependable object detection model.
mAP (Mean Average Precision): The uprising trend in the mAP values, both at an IOU threshold of 0.5 and across multiple IOU thresholds ranging from 0.5 to 0. The plot indicates that the model's predictions are collaborating well with ground truth annotations as training progresses regardless of different threshold values. This is an important indicator of the overall ability of the model in object detection tasks. These trends overall indicate that the YOLOv8 model is learning well from training data and advancing its capacity for image object localization and classification. The fact that these metrics continue to improve means the model is becoming more effective at refining its understanding of visual features associated with different object classes, as well as getting better and able to make accurate predictions. The consistent improvement in all these metrics indicates that the model is refining its

understanding of the visual features associated with different object classes and is becoming more adept at making accurate predictions.

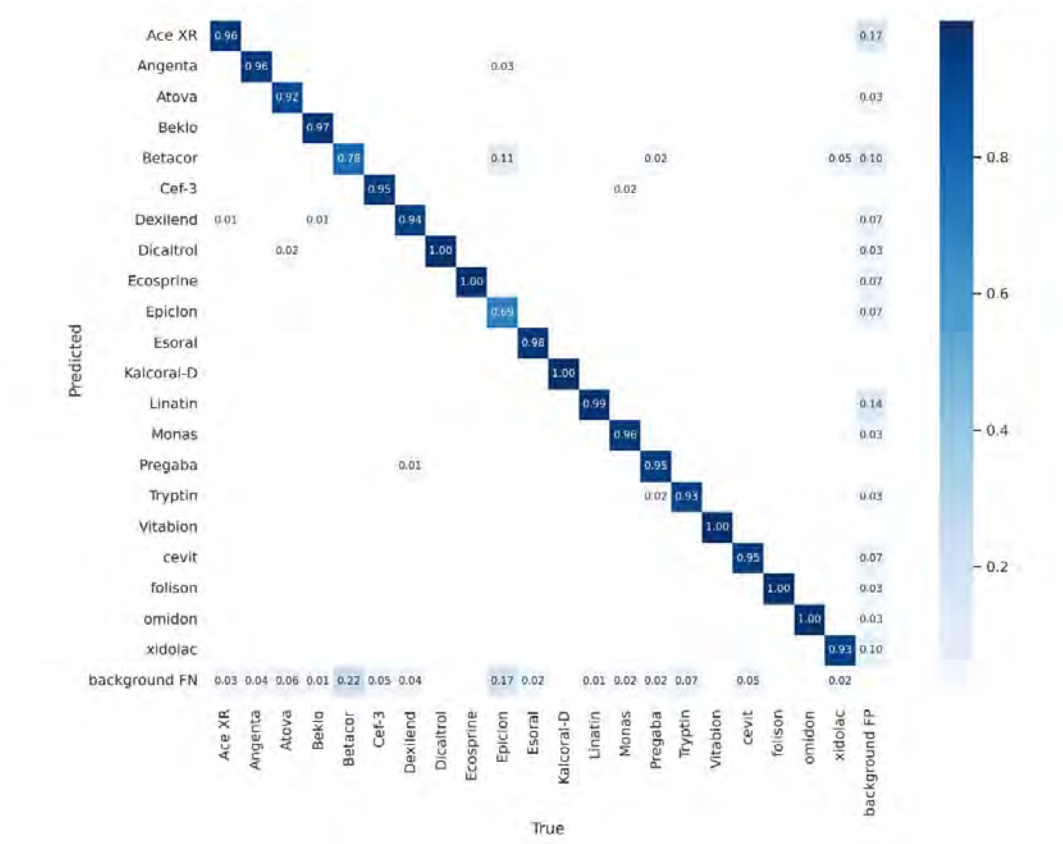


Figure 4.14: Normalized Confusion Matrix for YOLOv8.

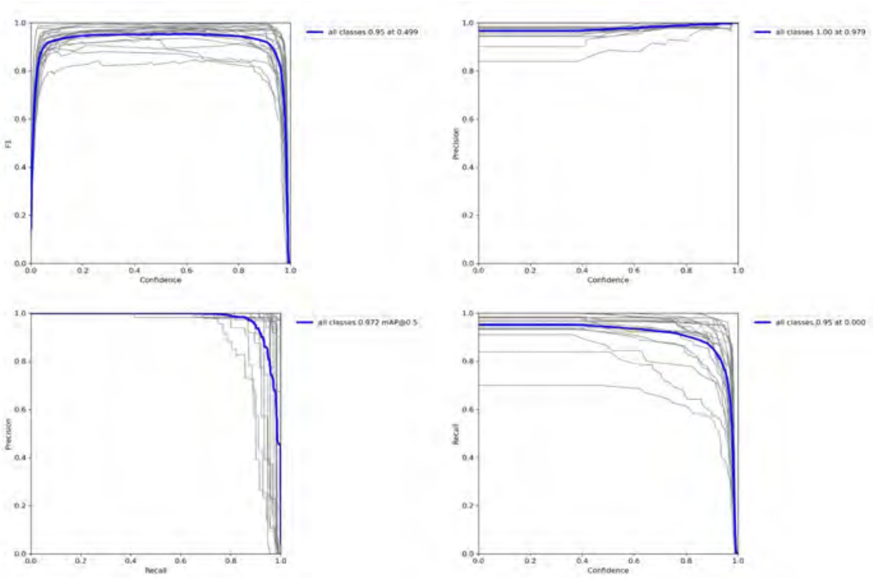


Figure 4.15: Performance Metrics for YOLOv8.

In our study, we assessed various metrics to evaluate the performance of YOLOv8. Our results show a huge precision rate of 97%, directing the model's ability to precisely identify relevant objects within the images. Moreover, our model gained a recall of 92%, improvising its proficiency in capturing a significant portion of the actual positive instances. In addition, the mean average precision at an intersection over the Union IOU threshold of 0.5 was found to be estimated as high and its value is performed for this model in terms of both precision and recall at this threshold. Also, the mAP at IoU thresholds ranging from 0.5 to 0.95 was found as that of 0.93 which signifies an average performability by model for distinct ranges in IoU threshold values. These findings reinforce the YOLOv8 model's high performance in object detection tasks for use cases applicable to real-world scenarios.

Figure 4.16: Labeled Image (YOLOv8).

The YOLOv8 model shows significantly better performance in learning to detect and classify backside strips. As shown by both the loss curves and performance metrics we can see that it reflects positive outcomes. Notably, the model demonstrates a high precision which means that most of the predicted strips are true positives. Furthermore, the model shows an improved recall score compared to YOLOv7.The results suggest a reduced chance of missing actual backside strips while detecting from images. Another thing to mention is that YOLOv8 surpasses in handling scenarios when we show multiple images in a single frame. Because of this, we could find an approach to tackle the challenges experienced by YOLOv7. These statistically improved results stand for better accuracy and a notable reduction in detection failures. Furthermore, It highlights the model's better capability to accurately detect and classify backside strips when we show it multiple medicines at once.

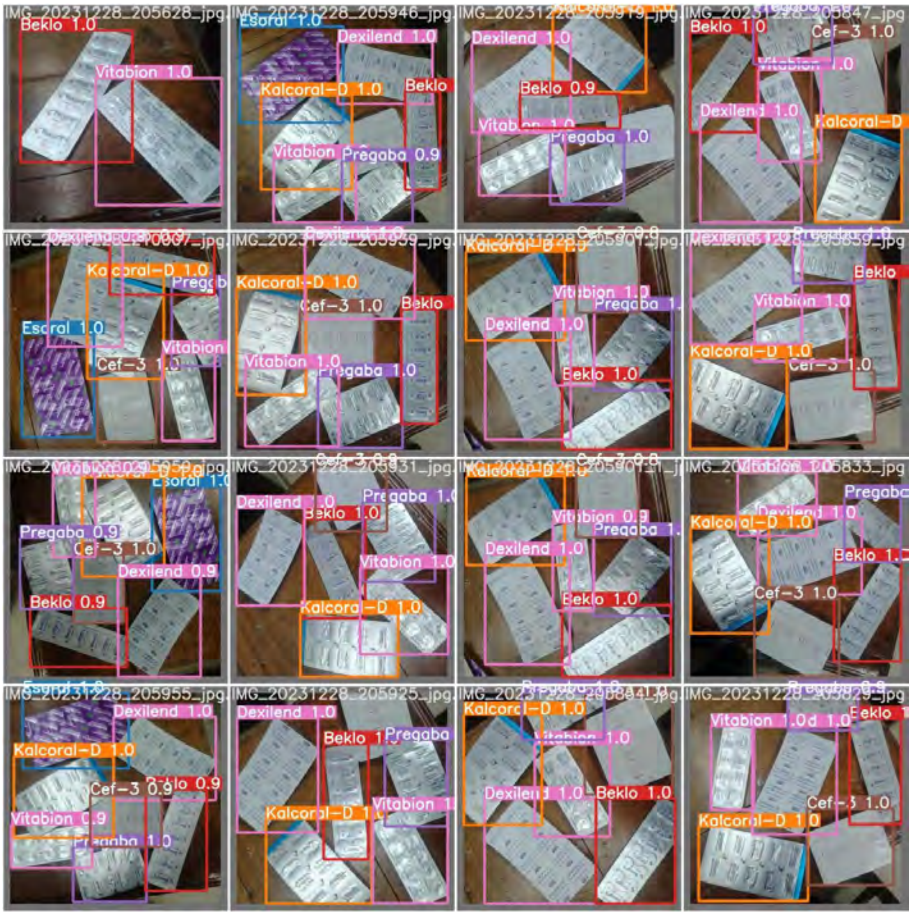


Figure 4.17: Predicted Image (YOLOv8).

4.1.12 Result Analysis

Clear differences in the metrics of performance between YOLOv7 and YOLOv8 object detection tests indicate superior model architecture design as well as training practices. YOLOv8 shows notable improvements in some categories. 0.97 indicates that the 97% confidence level of YOLOv8 is far better precision compared to that of YOLOv7 having only 89.5% confidence level, meaning it can more accurately and effectively identify positive cases with reduced instances of false positives. Furthermore, YOLOv8 has a fair recall of 92% which means that the model can recover about half to three-quarters of relevant positives although not as fully effective compared to those in YOLOv7 with its higher recall at around 79.8%. YOLOv8 outperforms YOLOv7 with mean average precision(mAP) on mAP at 50 and mAP at 50-95, with percentages of 95% and 93%, respectively, compared to the YOLOv7's respective metrics of 87.6% and 83.6%, further confirming the superiority at object detection capabilities over its predecessor, YOLOv7.

Model	Precision	Recall	mAP@50	mAP@50-95
YOLOv7	0.90	0.80	0.88	0.84
YOLOv8	0.97	0.92	0.95	0.93

Table 4.3: Comparison of Results Table.

4.1.13 Real-time detection Results

While we tested our system in real-time with a webcam, our used model of YOLOv8 showed amazing performance while detecting the back side label of the medicine strips. When correctly identifying intact medicine backstrips, the model often achieved a high confidence level of around 95%. Additionally, the model showed reliable detection with confidence scores of 85% to 90% even when a backstrip was broken after removing the tablet in case it failed. The model was also able to show its adaptability by correctly identifying clipped strips at about 85% confidence level. These results show that YOLOv8 is flexible in practical settings, especially when

Figure 4.18: Realtime Test Results.

the detection involves situations involving medication where perfect identification of strips containing medications matters a lot. The model's ability to differentiate between cut strips and damaged ones highlights its effectiveness for application in everyday life.

4.2 Fall Detection

The process commences with real-time camera monitoring for fall detection. The camera monitors video in real-time, enabling continuous detection of fall patterns using trained data. This includes various fall scenarios, encompassing both accidental and intentional instances. Upon detecting a fall, a signal is promptly generated and conveyed through a speaker system, announcing "Fall Detected" to alert nearby individuals for immediate assistance. Simultaneously, an email alert is sent to the designated caretaker, ensuring timely notification and necessary intervention.

4.2.1 Mediapipe

We attempted Mediapipe which is an open-source framework by Google. Mediapipe's pose detection model provides key landmarks for each body part and these landmarks are then tracked using OpenCV's optical flow algorithm to estimate the motion of the landmarks. When processing a 2D image or a video frame, Mediapipe identifies and tracks 2D landmarks, such as joints and key points. These 2D landmarks serve as reference points. Mediapipe has the knowledge of human body structure and pose estimation algorithms which helps it to figure out the 3D position of those landmarks. Through this analysis, Mediapipe can give us a 3D vector representation of a 2D image.

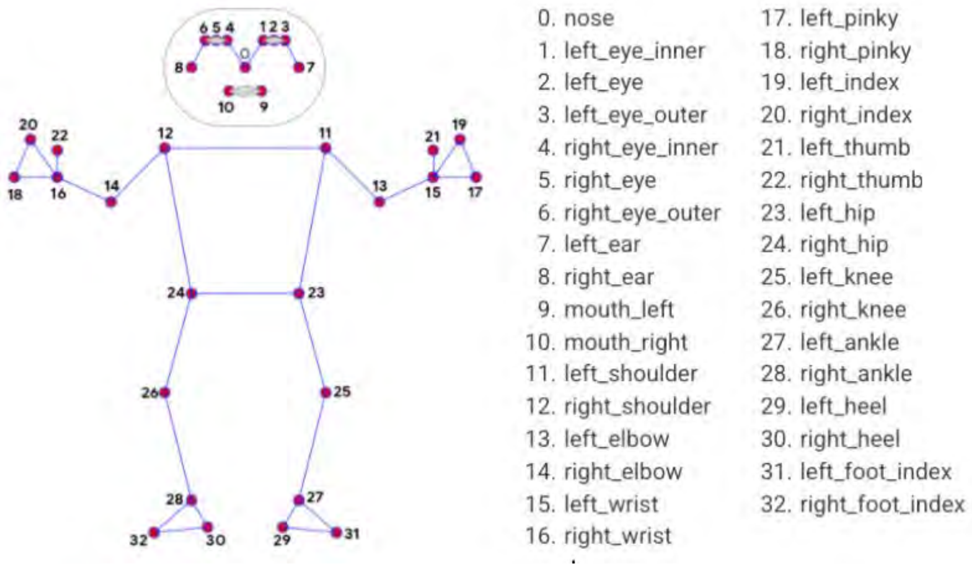


Figure 4.19: Mediapipe landmarks [24] .

The model we used for fall detection using MediaPipe was ultimately deemed non-viable for several reasons. Despite its initial promise, the use of key points on the body as landmarks did not adequately capture the necessary patterns for fall detection. Human falls can manifest in a wide variety of patterns and our model's reliance on specific body landmarks proved insufficient for accurately detecting these diverse patterns. As a result, we decided to take a different approach to effectively and reliably detect falls in real-world scenarios.

4.2.2 YOLOv8

To overcome the challenges, we encountered with our initial model, we made the decision to switch to YOLOv8 for a different approach to fall detection. YOLOv8, a fine model for object detection, provided a better solution to overcome the shortcomings of our previous method for identifying and analyzing fall-related patterns.

By using the advanced capabilities of YOLOv8, we aimed to strengthen the accuracy and relevance of our fall detection system. It enabled us to address the complex nature of human falls more effectively.

4.2.3 Dataset

The dataset used for our fall detection model was sourced from Roboflow. It has a wide range of fall patterns captured in 10,793 images. These images were taken carefully to have diverse scenarios of falls, providing a comprehensive sample of real-world fall patterns. By using a rich and varied dataset, we aimed to ensure that our model was exposed to a lot of fall-related movements and postures. That enabled the model to learn and recognize different types of falls. This approach significantly improved the model and contributed to becoming more useful in real-world applications.

Figure 4.20: Various Fall Patterns [19] .

4.2.4 Train Valid Test Split

We split the dataset into three subsets for training, validation and testing. The training set was 88% of the data, totaling 9,444 images. The validation set consisted of 8% of the data, comprising 899 images, while the remaining 4% of the dataset, which was 450 images, was used as the test set. The lower number of images in the test set was our choice, reflecting the intent to prioritize manual real time testing. This strategy aimed to ensure that the model's performance could be thoroughly evaluated in real-world settings.

4.2.5 Training and Results

The model was trained using the Adam optimizer, a popular choice for optimizing deep learning models. This model is known to be the best for deep learning models. The training process was 300 epochs with a batch size of 64, as it was considered essential to expose the model to numerous iterations. This made the model capable of capturing complex patterns associated with falls. A learning rate of 0.001 was utilized to adjust the model's weights during training, which struck a balance between the need for steady convergence and the risk of overshooting minima.

Figure 4.21: Sample Batch of Annotated Images for Fall Detection Training.

During the training phase, the model exhibited a consistent decrease in both box, loss and class loss. This indicated the improvement in its ability to predict bounding boxes and classify objects correctly. The total loss, which combines various components of loss, also showed a decreasing trend, reflecting overall enhancement in the model's capabilities. Validation results showed the positive trends observed in the training phase. With valida- tion we saw box loss, class loss and total loss. All of them indicated a decline. This suggests that the model was generalizing well and not overfitting to the training data. Furthermore, the model's precision, recall and mean Average Precision (mAP) at IoU thresholds of 0.5 and 0.95 were consistently high and stable during the training process. These metrics indicate that the model was adept at accurately identifying fall-related objects. It had high precision in its predictions and demonstrated robust performance in terms of both recall and overall precision.
The confusion matrix presented in the research paper report provides a comprehensive overview of the classification model's performance. It particularly distinguishes between "Fall-Detected" and "background" classes. The matrix illustrates the distribution of true positive, false positive, true negative and false negative predictions. With a value of approximately 0.89 for true positives in the "Fall-Detected" class,

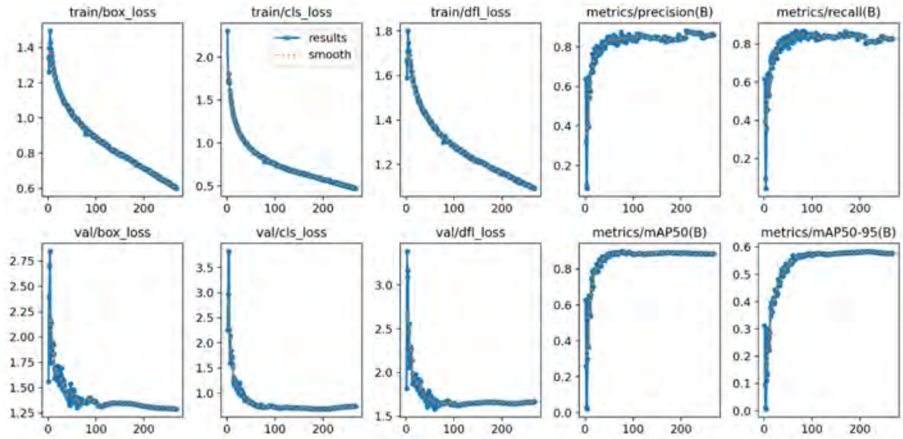


Figure 4.22: Training results for YOLOv8 Fall detection model

the model shows a high rate of correctly identified falls. However, the presence of false negatives at around 0.11 suggests instances where falls were not detected, indicating a potential area for model improvement. Conversely, the model shows satisfactory performance in correctly identifying instances of the "background" class, as evidenced by a normalized value of 1.00 for true negatives. Overall, the confusion matrix serves as a crucial tool for evaluating the model's classification performance and highlights key areas for further optimization.

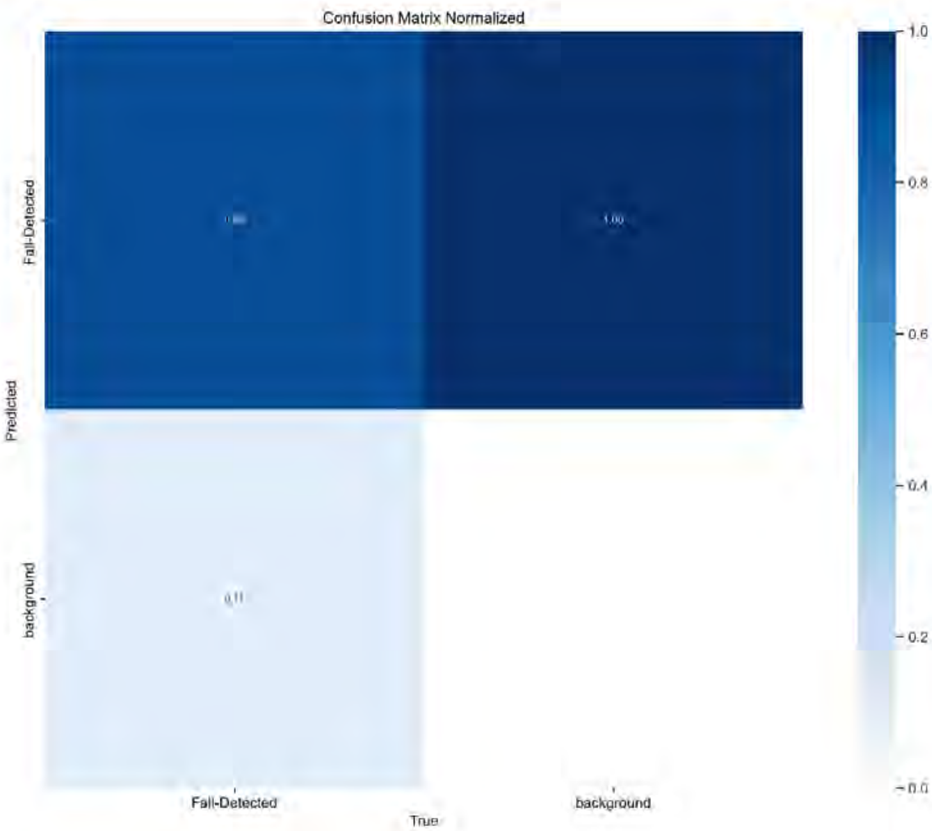


Figure 4.23: Normalized Confusion Matrix

The validation result shows the effectiveness of the model in identifying and clas-

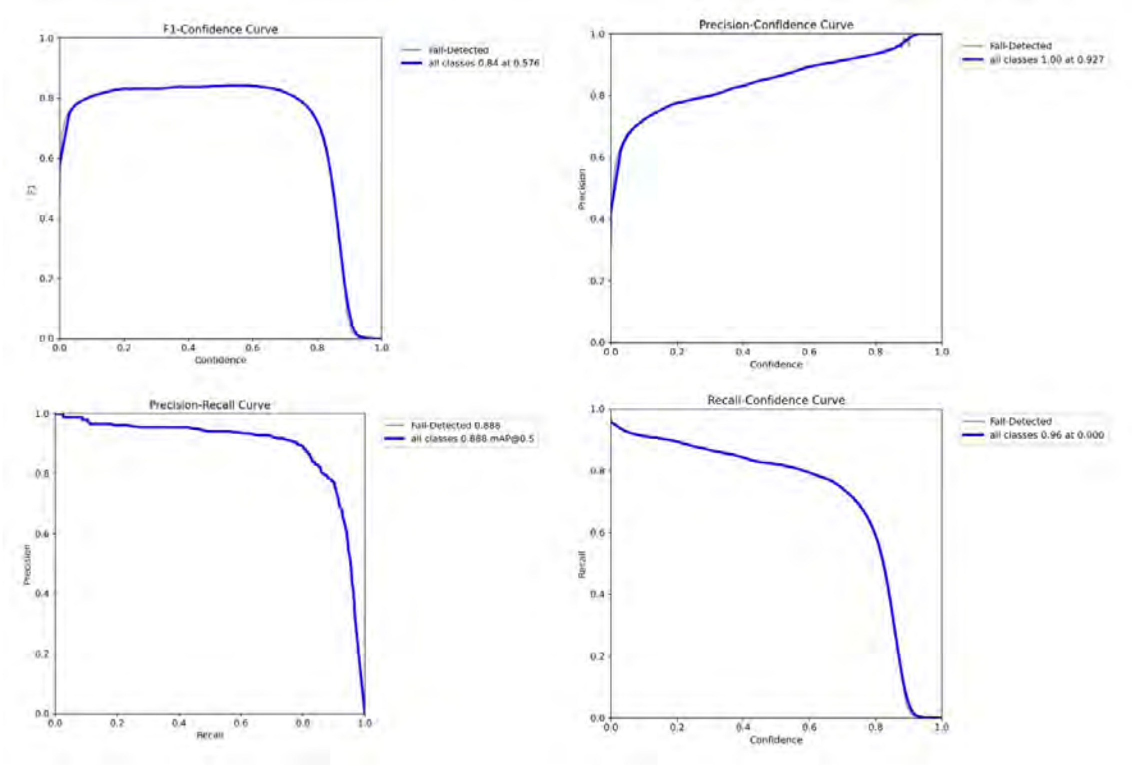


Figure 4.24: Performance Metrics for Fall Detection Model

sifying objects of interest. The precision value of 0.874 indicates that when the model predicts an object as belonging to class B, it is correct approximately 87.4% of the time. This high precision value suggests that the model has a low rate of false positives, making it reliable. Furthermore, the recall value of 0.812 indicates that the model successfully identifies around 81.2% of all actual class B instances. This indicates that the model has a relatively low rate of false negatives. The mean Average Precision (mAP) at a threshold of 0.5 (mAP50) further underscores the model's strong performance, with a value of 0.888. This metric provides an overall assessment of the model's precision-recall trade-off. This indicates a high level of accuracy in object detection and classification at the specified threshold. Additionally, the mAP50-95 value of 0.585, which considers a stricter IoU threshold of 0.95. This reflects the model's performance in accurately localizing and classifying objects that are overlapping. While the mAP50-95 value is lower than the mAP50, this is expected due to the more stringent evaluation criteria. The achieved value still indicates a commendable performance in object localization and classification.

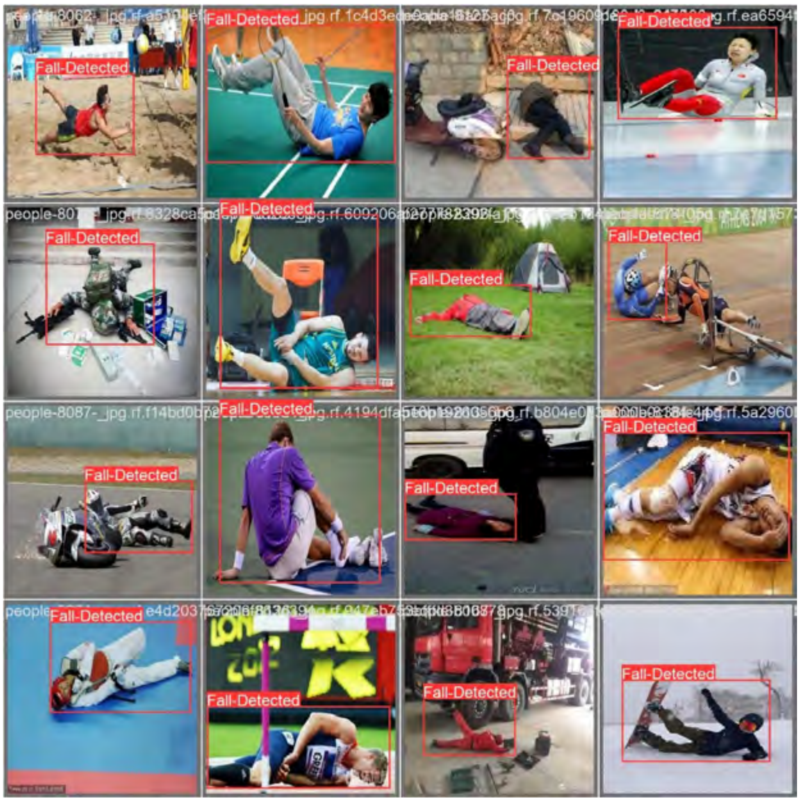


Figure 4.25: Sample Batch of Labeled Images for Fall Detection

Figure 4.26: Sample Batch of Predicted Images for Fall Detection

4.2.6 Real-time detection Results

The real-time testing of fall detection using a webcam in a home environment yielded promising results, with the model consistently achieving high confidence levels ranging from 0.80 to 0.90 in accurately detecting falls. For optimal performance of our fall detection model, the person's full body must remain visible, as obscured or partial visibility may lead to inaccurate results. This performance underscores the model's identifying critical events within a domestic setting, offering a reliable means of ensuring prompt assistance in the event of a fall. The model we used was able to demonstrate the power of using a webcam's abilities regarding fast and confident detection and classification of the fall mechanism that would increase safety and comfort within the home. Real-time testing in circumstances outside of the lab setting also allowed for real-world feasibility to be widely confirmed, further reinforcing that this model would not only have effective results due to its ability. However, it has the capacity as a tool for proactive fall detection and intervention within residential care homes.

Figure 4.27: Realtime Fall detection scenarios

4.2.7 Sending Alert

This module uses the "smtplib" library so that it can connect to the Gmail SMTP server to send email notifications. This process helps the code to set up a secure connection to the SMTP server by sharing the sender's information and settings for logging into their email account so that it can automatically send alert emails to senders who are set before to receive each time after noticing any fall. On top of the email notification mode, a code can be integrated to include alerts in voice communications. This can be achieved by introducing a speech synthesis API or TTS library as a means of producing an audio alert whenever falling is detected. With the interface of the YOLO model that detects a fall, there is a code to run TTS sensitivity while allowing it to pick up on falls and receive alerts verbally alerting those nearby about this issue. This function also ensures the caregivers of the elderly receive emails and everyone nearby hears a loud alert that the elderly have fallen. Combine the emailed fall alerts and audio beeps so responding to falls quicker and more completely. This integrated scheme utilizes voice and email chains to report falls effectively with quick intervention.

4.3 Voice Assistant

The voice assistant that we developed is adjusted to natural human language understanding and it can easily handle simple question-answer tasks. Furthermore, it also incorporates some state-of-the-art technologies to give the best conversational experience possible for our elderly citizens. The implementation of this system focuses on unstructured questions-answering, profiting from the unique abilities of Large Language Models(LLMs), together with the innovative development technology called Retrieval Augmented Generation(RAG). We used some special tools such as TextLoader and Carefully divided the text with the help of TextSplitters. These are used to feed the fundamental architecture associated with intelligence within the voice assistant system. Furthermore, It can take in an incoming stream of text data. This process makes sure that the process is efficient, enhancing models to better understand the complex questions. Each text fragment is first sent through a transformation and converted into a numerical representation through the OpenAI Embeddings and then made to be stored in the Chroma Vector store. This was done by VectorstoreIndex Creator. However, chapter three talks about introductions of new organization formations and operating systems. While initiating a conversation, this smart structure enhances the storage of data and speeds up its retrieval as well. These tools and frameworks culminate in a powerful conversational agent, emphasizing LLMs, RAG and the robust LangChain framework. The result is an intelligent and adaptive companion designed for elderly individuals, excelling in comprehending and responding to their queries, with a sincere focus on unstructured data interpretation.

4.3.1 Retrieval Augmented Generation Architecture (RAG)

The heart of our voice assistant lies in the implementation of Retrieval Augmented Generation, a technique designed to enhance the knowledge base of Large Language Models. RAG facilitates the reasoning over both public and private data, overcoming the limitations imposed by the model's cutoff date. The process involves two main components:

- **Indexing:**The system ingests data from a source, breaks it into smaller chunks for efficient processing and indexes these chunks. This offline process prepares the data for quick retrieval during runtime. We employed DocumentLoaders for data loading, Text Splitters for breaking large documents into manageable chunks and a VectorStore combined with embeddings for efficient storage and indexing.

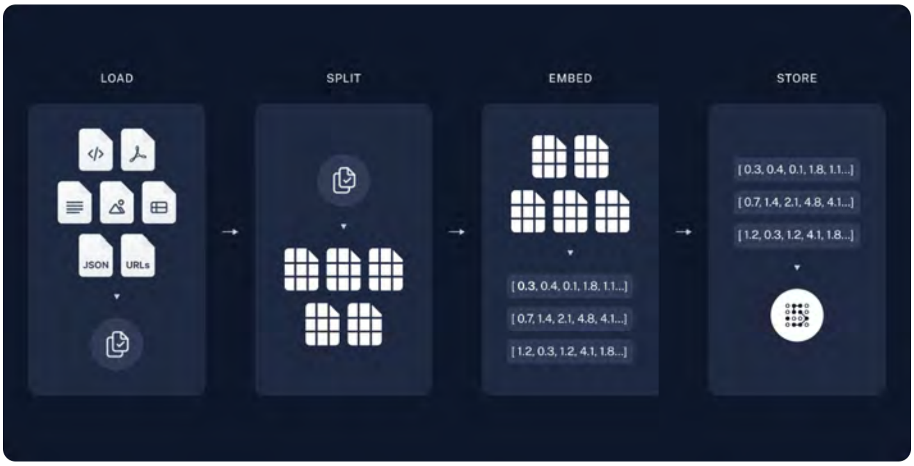
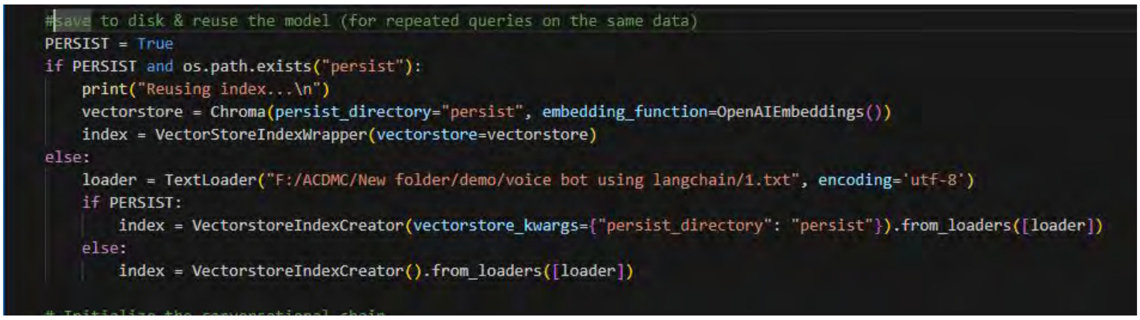


Figure 4.28: Indexing using Retrieval Augmented Generation [17]



```
#save to disk & reuse the model (for repeated queries on the same data)
PERSIST = True
if PERSIST and os.path.exists("persist"):
    print("Reusing index...\n")
    vectorstore = Chroma(persist_directory="persist", embedding_function=OpenAIEmbeddings())
    index = VectorStoreIndexWrapper(vectorstore=vectorstore)
else:
    loader = TextLoader("F:/ACDMC/New folder/demo/voice bot using langchain/1.txt", encoding='utf-8')
    if PERSIST:
        index = VectorstoreIndexCreator(vectorstore_kwargs={"persist_directory": "persist"}).from_loaders([loader])
    else:
        index = VectorstoreIndexCreator().from_loaders([loader])
```

Figure 4.29: Example code for Retrieval

- **Retrieval and Generation:** During runtime, the RAG chain retrieves relevant data from the index based on user queries and passes it to the language model for answer generation. The retriever component identifies the most pertinent information, while the ChatModel, utilizing the OpenAI GPT-3.5 Turbo model, produces a coherent and contextually relevant answer.

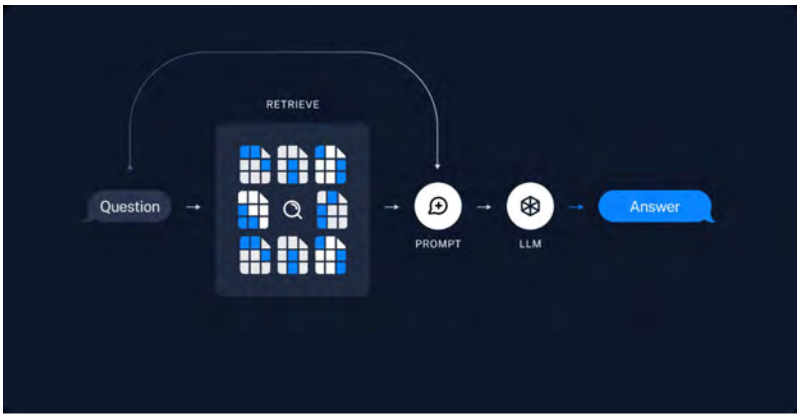


Figure 4.30: Retrieval and Generation using Retrieval Augmented Generation [17] .

4.3.2 Conversational Chain

The voice assistant utilizes the Conversational Retrieval Chain, incorporating the [26] ChatOpenAI model for language understanding. ConversationalRetrievalChain in LangChain seamlessly bridges the gap between the LLM and the retriever. This chain manages the flow of conversation, integrating both user input and the model's responses into a coherent dialogue. The chat history is preserved, allowing for context-aware responses and improved user interaction. This chain maintains a record of your chat history, enabling the bot to understand the context of the conversation and generate increasingly relevant responses as the dialogue progresses. It's like having a friend who remembers everything you've said, leading to natural and engaging conversations.

```
# Initialize the conversational chain
chain = ConversationalRetrievalChain.from_llm(
    llm=ChatOpenAI(model="gpt-3.5-turbo"),
    retriever=index.vectorstore.as_retriever(search_kwargs={"k": 1}),
)
```

Figure 4.31: Example code for Conversational chain

4.3.3 User Interaction Loop

A continuous loop system that fosters ongoing and supportive interactions, particularly tailored for the elderly through the creation of our voice bot. This empathetic conversational AI engages in friendly dialogue, aiming to alleviate loneliness and provide companionship. By prioritizing natural voice interaction, empathetic dialogue and human-like responses, we have successfully transformed a conventional conversational AI system into a voice bot dedicated to addressing the unique needs of the elderly population. Within this system, the main loop orchestrates a seamless interaction process, where the assistant, equipped to process user queries, retrieve relevant information and maintain conversation history, ensures a context-aware and human-like interaction, contributing positively to the well-being of elderly users.

4.3.4 Voice Input and Output Handling

Our voice assistant integrates the [25] Google Speech Recognition API for capturing user voice input. The system prompts users with spoken questions and processes their responses through the speech recognition module. Text responses are then converted back into speech using the gTTS (Google Text-to-Speech) library, providing a natural and immersive conversational experience in the Bengali and English languages. Our implementation includes audio playback functionality using the Pygame and [9] pyttsx3 library. The assistant saves generated audio files, facilitating the seamless playback of responses.

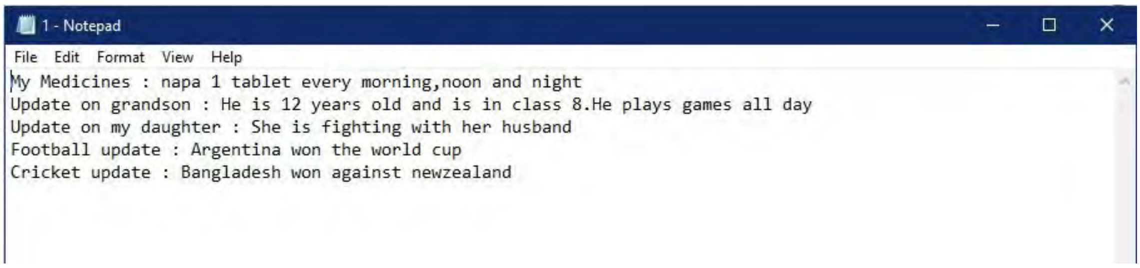


Figure 4.32: Example Information for Retrieval

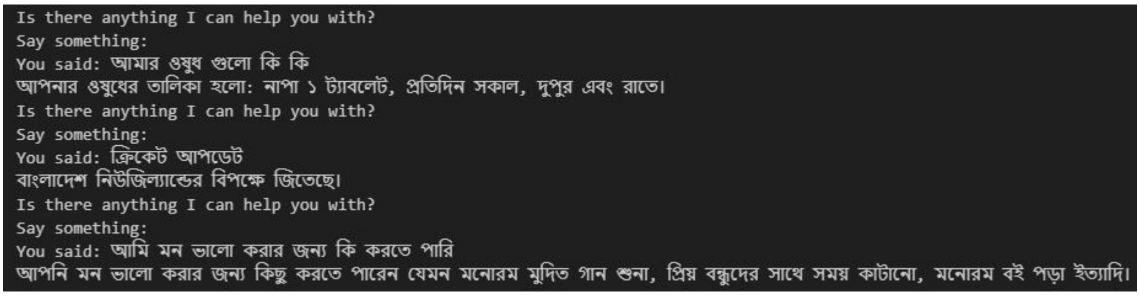


Figure 4.33: Example Conversation

We can not only retrieve information from preloaded datasets using advanced data indexing and retrieval techniques but also seamlessly integrate OpenAI's Language Model (LLM) to generate responses when data is not readily available. The implementation utilizes the Conversational Retrieval Chain from LangChain, allowing the system to maintain a continuous interaction loop with users. When a user inquiry matches existing data, the bot efficiently retrieves and presents relevant information. However, what sets this voice bot apart is its ability to dynamically answer questions by leveraging the ChatOpenAI model, even when the required data is not explicitly present in the pre-loaded document. This incorporation of the LLM ensures a more versatile and contextually aware conversational experience, making the voice bot adept at handling a wide array of queries and providing informative responses, further enhancing its utility and effectiveness.

4.4 Voice Activated Emergency Assistance

This feature focuses on utilizing speech recognition technology to enable users to send emergency email notifications through voice commands. The core of the system lies in the integration of the SpeechRecognition library, which allows the system to transcribe spoken language into text. Leveraging this technology, users can trigger an emergency email request by uttering specific phrases.

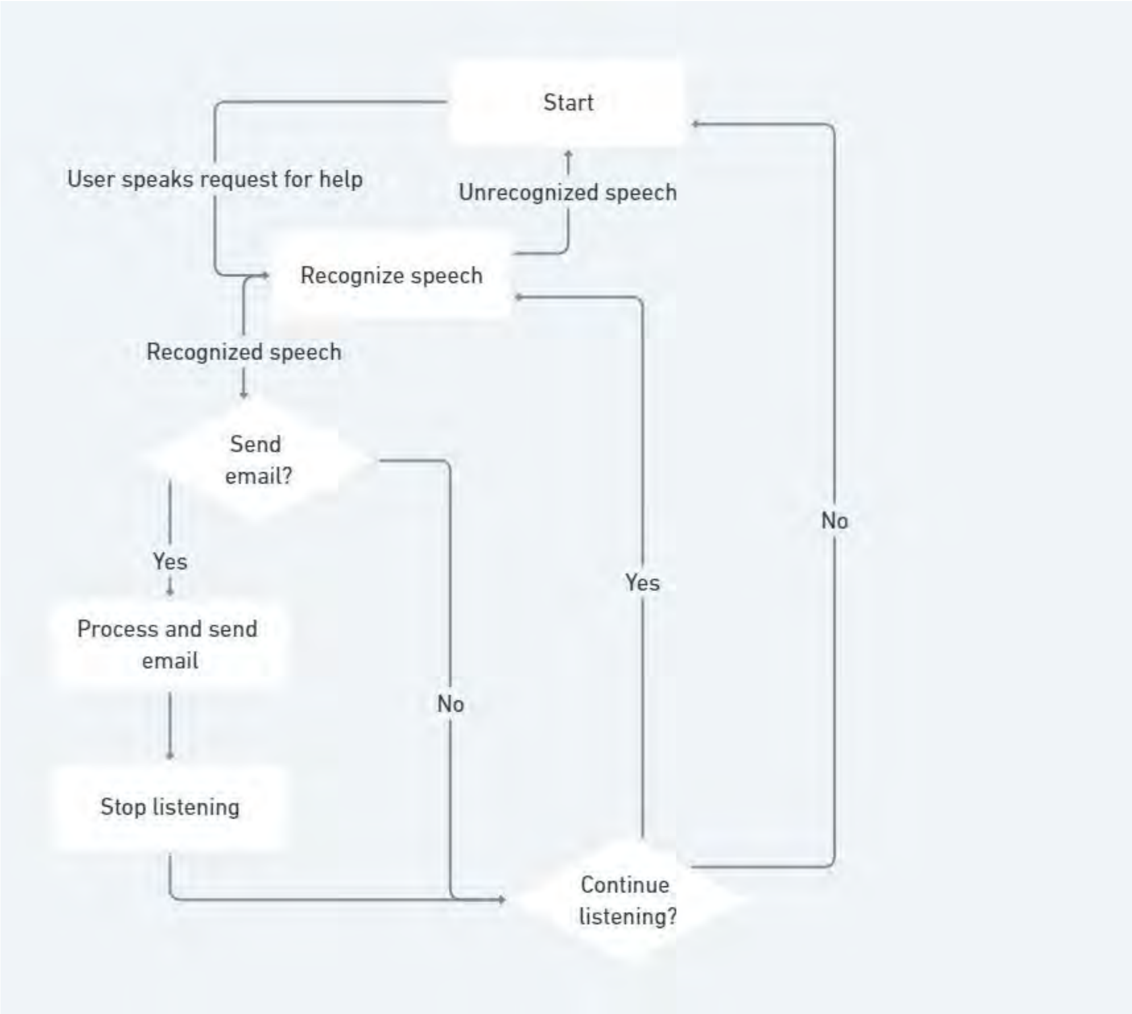


Figure 4.34: Emergency Assistance Process

4.4.1 Speech Recognition Module

The recognize_speech function utilizes the [23] Speech Recognition to transcribe the spoken words into text. The system employs ambient noise adjustment to enhance accuracy, ensuring robust performance even in varied acoustic environments. In case of an unrecognized speech or a request error, the system gracefully handles these scenarios and returns an empty string. The implementation also operates within a continuous loop, perpetually listening for user input. This loop ensures the real-time responsiveness of the system to potential emergency requests. Additionally, an optional condition is integrated to break the loop based on a user-specified phrase.

4.4.2 Email Notification Functionality

Upon recognizing the trigger phrase in the spoken text, the system seamlessly initiates the send_email function. This function is responsible for constructing an email with a predefined subject and body, expressing the need for emergency help. The email is sent from a specific sender to a designated caretaker via the Gmail SMTP server. Security is maintained through password-protected authentication during the login process. This feature offers a practical and hands-free solution for elderly individuals who may find it challenging to manually type or dial for emergency assistance. By combining speech recognition and email functionalities, our system provides a quick and efficient means for users to seek help, potentially enhancing personal safety in critical situations. The robustness of the implementation is underscored by its ability to adapt to different acoustic conditions and gracefully handle errors, making it a valuable addition to voice activated emergency response systems.

4.5 Reminder

We implemented a modern and very basic reminder feature for better ease of access. A reminder system is a crucial feature for any elderly person. Our research is primarily focused on elderly citizens of our community. Thus, a very common problem they face is remembering things. They often forget crucial events, meetings, or work. Many elderly people forget prayer time and regret it afterward. But the most important one is that they often forget to take their medicines. We conducted a survey among 100 elderly people throughout the community we live in. We found that approximately 75% of them can not remember the tasks they are supposed to complete in their day-to-day life. Bearing these in mind, we built a reminder system which will help the elderly conquer all these challenges. Our reminder system can notify the elderly in both English and Bangladeshi languages. More language support will also be available in future updates.

4.5.1 Functionalities

The basic functionalities of our reminder system are as follows;

- Remind medicine intake times
- Remind correct medicine names
- Remind important meeting times
- Alert for prayer times
- Can be used to set an alarm

```python
import time
from gtts import gTTS
import pygame
import os
```

Figure 4.35: Used libraries

4.5.2 Used Libraries

Our system is created using python language as its base. The reminder system is also written in Python language. We used the following libraries to facilitate the reminder system;

- **Time:** The time library enables us to get the current time in seconds. It can also be used to stop the execution of the code for a particular amount of time. Again, we used this library to format the current UTC/ global time and to convert it into local time. Additionally, time library is used to fetch the local time for better synchronization. We also used 'time' to measure the elapsed time.

- **gTTS:** 'gTTS' means "Google Text to Speech" in short. It's a Python tool and helps people talk with Google Translate's speech system for text through simple commands. The main job of 'gTTS' is to change written words into spoken ones, making voice from the given text input. It uses Google's Text-to-Speech API to make an audio file that matches the text. Google Translate works with a lot of different languages and gTTS lets us pick the language for the text being changed. This makes our reminder adaptable for creating voice material in different languages.

- **Pygame:**In terms of our reminder system, we used the Pygame library for audio playback. In simple terms, it is used to play the created voice messages as MP3 files. The code starts with setting up the Pygame mixer using pygame.mixer.init(). This step starts the Pygame mixer, which is a tool that helps to play sounds from files. It has to be prepared before we can use any sounds-related tasks. In the play reminder function, pygame.mixer is used to load and start playing an MP3 file that was just made. After a certain time (20 seconds in this case), the code stops playing sound. Before deleting the old MP3 file and saving a new one, the code gets rid of music. 'pygame.mixer.music.unload()' this line gets rid of the music, freeing up its file handle for other use. When the code is over, pygame mixer stops. In short, pygame helps with controlling and playing sound files in the code. It makes it easy to handle playing back sounds using Python. This is good for situations when we just need basic audio features, like hearing voice alerts set for taking medicine in our reminder system.

- **OS Module:**We used the 'os module' for handling files and folders. It is used to make complete file paths, check if files are there, remove old sound files before saving new ones and create directories for audio stuff if they don't exist. It can also find out where we currently work on our computer. These tasks are important to manage the sound files made by our script. They help keep file paths right, stop too many files from building up and organize storing voice notices.

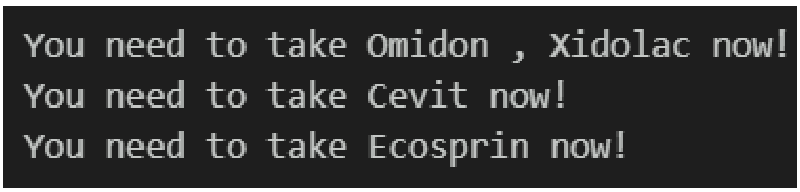


Figure 4.36: Snap of the code output

Chapter 5

Conclusion

5.1 Conclusion

In conclusion, the development of a bot system designed specifically for elderly individuals, capable of detecting common medications, reminding them to take their appropriate doses, monitoring their posture for potential dangers, and engaging in interactive conversations, holds tremendous promise for enhancing the over-all quality of life and well-being of the elderly population. The integration of advanced technologies such as artificial intelligence, real-time image processing, and computer vision, has paved the way for the creation of intelligent robots that can address the needs and challenges faced by elderly individuals. While the development of a system for elderly care represents a significant advancement in the field, there are several ways in which it can be further improved. Providing the bot system with more data and information will result in increased accuracy and improvement over time. Therefore, it has a high potential for further improvement. With further research, this technology can make a significant and positive impact on the lives of the elderly. Therefore, the implementation of a system that can detect medications, provide reminders, monitor dangerous postures, and engage in conversations has the potential to revolutionize elderly care.

5.2 Future Scope

We initially started with a vision to improve the daily lives of elderly people. Eventually, we realized that the boundaries of our research are huge. We can contribute to various sectors of our society using our research. But due to many limitations we only focused on some selected sectors of elderly people's lives.

Our first and foremost concern was detecting medicines accurately. We initially started with 21 medicines to detect for our elderly. In the context of our country, there are no available datasets regarding medicines. Thus, we used the distinct labels of the medicines to detect the medicines. To do this, we created our very own primary dataset with a limited number of data. If we can improve our dataset with more data and increase the number of medicines covered then our system will definitely perform a lot more accurately.

Furthermore, it is very difficult to detect medicines using their shape or color because in our country they do not use any unique features in a tablet which can be introduced as distinct. So, in the near future, if the medicine companies introduce

some unique code on the tablets and medicines, it would be very useful for our system to detect medicines using color and shapes which will further improve our system's accuracy.
We can incorporate functionality for keeping track of medicine intakes and the remaining inventory of our consumers in the near future. We can also send notifications to the caregiver when the inventory becomes low so that the elderly are never short on crucial medicines.
If implemented in a robust bot, our system can further have various sensor information and use them to monitor the elderly further. For example, our system would be able to measure blood pressure, blood sugar, body temperature, oxygen level and weight of the elderly. This would provide more comprehensive care.
Our system may keep track of the eating habits and calorie intake to ensure a healthy diet for elderly people. Furthermore, it may also remind the elderly to do regular exercise and monitor their exercises.
A new voice control feature may be implemented in the future which would enable the user to adapt with home appliances and turn the elderly people's home into a smart home. For instance, our bot system may control the temperature of the air conditioner, room heater, TV, fan or lights etc through the user's voice. This would ensure the optimization of resources as well as make the elderly lives comfortable and safe.
We may introduce a new and more friendly interface and technology in the future by which the elderly would be able to behave more friendly towards our system. For example, our system might detect human emotions and interact with the user accordingly. Also, our system would be able to behave in a more human-like way so that the elderly do not feel left out or lonely.
Additionally, our system might incorporate some sort of IR technology which would allow the user to seek help from the caregiver in case he/she needs help to set up or troubleshoot a modern device from another location. i.e; Smartphone, Smart TV, Laptop, Router etc.
Furthermore, our system may keep track of the daily habits of our elderly people to better monitor their lifestyle and send this information to the caregiver who will monitor the elderly. This information may further be used to improve their daily lives and instruct them to stay healthy by building healthy habits.
Our system initially works with only a single user at a time. But we can make it work for multiple users in the near future. That way our system might be an effective solution for old age homes.
We might incorporate a feature which will detect more postures in the future. We can detect prayer postures, ideal sleeping postures, sitting postures etc. Also, we might introduce more dangerous postures in the future so that we can prevent any dangerous situation which may risk the well-being of the elderly.
We can also ensure the posture of medicine intake to detect if the user is taking the medicine timely or not and inform the caregiver. The system will be able to inform the user of the medicines they take and what are the applications of those medicines.
In our survey, we found that elderly people need various important items on a daily basis. We can introduce those items to our system so that our system can detect those items within its field of view. That way it would be more tangible for the elderly to use our system.

We can further conduct studies on a larger number of elderly people and let them use our system. By conducting further studies, we can discover more crucial needs of the elderly people and introduce them to our system.

Bibliography

[1] M. Yu, A. Rhuma, S. M. Naqvi, L. Wang, and J. Chambers, "A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, pp. 1274–1286, Nov. 2012. DOI: 10.1109/titb.2012.2214786. [Online]. Available: https://ieeexplore.ieee.org/document/6279483.

[2] J. Woo, J. Botzheim, and N. Kubota, "Conversation system for natural communication with robot partner," in *2014 10th France-Japan/ 8th Europe-Asia Congress on Mecatronics (MECATRONICS2014- Tokyo)*, 2014, pp. 349–354. DOI: 10.1109/MECATRONICS.2014.7018624.

[3] F. Harrou, N. Zerrouki, Y. Sun, and A. Houacine, "Vision-based fall detection system for improving safety of elderly people," *IEEE Instrumentation & Measurement Magazine*, vol. 20, no. 6, pp. 49–55, 2017. DOI: 10.1109/MIM.2017.8121952.

[4] H. G. Ng, P. Anton, M. Brügger, *et al.*, "Hey robot, why don't you talk to me?" In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2017, pp. 728–731. DOI: 10.1109/ROMAN.2017.8172383.

[5] C.-Y. Yang, M.-J. Lu, S.-H. Tseng, and L.-C. Fu, "A companion robot for daily care of elders based on homeostasis," in *2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, 2017, pp. 1401–1406. DOI: 10.23919/SICE.2017.8105748.

[6] P. Joglekar and V. Kulkarni, "Humanoid robot as a companion for the senior citizens," in *2018 IEEE Punecon*, 2018, pp. 1–4. DOI: 10.1109/PUNECON.2018.8745399.

[7] P. Parvin, F. Paternò, and S. Chessa, "Anomaly detection in the elderly daily behavior," in *2018 14th International Conference on Intelligent Environments (IE)*, 2018, pp. 103–106. DOI: 10.1109/IE.2018.00025.

[8] *Medicine Recognition from Colors and Text*, Association for Computing Machinery, Jun. 2019, pp. 39–43. [Online]. Available: https://dl.acm.org/doi/10.1145/3338472.3338484.

[9] N. M. Bhat, *Pyttsx3: Text to speech (tts) library for python 2 and 3. works without internet connection or delay. supports multiple tts engines, including sapi5, nsss, and espeak.* PyPI, Jul. 2020. [Online]. Available: https://pypi.org/project/pyttsx3/.

[10] M. Bundele, H. Sharma, M. Gupta, and P. S. Sisodia, "An elderly fall detection system using depth images," in *2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, 2020, pp. 1–4. DOI: 10.1109/ICRAIE51050.2020.9358330.

[11] M. Mehala and D. J. Viji Gripsy, "Voice based medicine remainder alert application for elder people," *International Journal of Recent Technology and Engineering*, vol. 8, pp. 2284–2289, Mar. 2020. DOI: 10.35940/ijrte.f7731.038620. (visited on 01/02/2021).

[12] Z. Smith, *Live-in care agency finds 1 in 3 elderly people more lonely in wake of covid-19*, Elder, May 2020. [Online]. Available: https://www.elder.org/the-elder/survey-on-elderly-loneliness/.

[13] S. Subhash, P. N. Srivatsa, S. Siddesh, A. Ullas, and B. Santhosh, "Artificial intelligence-based voice assistant," in *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, 2020, pp. 593–596. DOI: 10.1109/WorldS450073.2020.9210344.

[14] X. Wang and K. Jia, "Human fall detection algorithm based on yolov3," in *2020 IEEE 5th International Conference on Image, Vision and Computing (ICIVC)*, 2020, pp. 50–54. DOI: 10.1109/ICIVC50857.2020.9177447.

[15] G. Chen and X. Duan, "Vision-based elderly fall detection algorithm for mobile robot," in *2021 IEEE 4th International Conference on Electronics Technology (ICET)*, 2021, pp. 1197–1202. DOI: 10.1109/ICET51757.2021.9450950.

[16] X. Li and W. Chen, "Fall recognition algorithm for the elderly based on home service robot," in *2021 International Conference on Electronics, Circuits and Information Engineering (ECIE)*, 2021, pp. 329–335. DOI: 10.1109/ECIE52353.2021.00074.

[17] H. Chase, *Langchain*, GitHub, Oct. 2022. [Online]. Available: https://github.com/langchain-ai/langchain.

[18] K. Murray, *Elderly loneliness*, Medium, Apr. 2022. [Online]. Available: https://medium.com/@kimmurray96/elderly-loneliness-f9e344bff600 (visited on 01/19/2024).

[19] Roboflow, *Fall detection object detection dataset (v4, resized640_aug3x-accurate) by roboflow universe projects*, Roboflow, Dec. 2022. [Online]. Available: https://universe.roboflow.com/roboflow-universe-projects/fall-detection-ca3o8/dataset/4.

[20] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *arXiv (Cornell University)*, Jul. 2022. DOI: 10.48550/arxiv.2207.02696.

[21] G. Jocher, A. Chaurasia, and J. Qiu, *Yolo by ultralytics*, GitHub, Jan. 2023. [Online]. Available: https://github.com/ultralytics/ultralytics.

[22] P. Malani, D. Singer, M. Kirch, *et al.*, *Trends in loneliness among older adults from 2018-2023*, www.healthyagingpoll.org, Mar. 2023. [Online]. Available: https://www.healthyagingpoll.org/reports-more/report/trends-loneliness-among-older-adults-2018-2023.

[23] A. Zhang (Uberi), *Speechrecognition: Library for performing speech recognition, with support for several engines and apis, online and offline.* PyPI, 2023. [Online]. Available: https://pypi.org/project/SpeechRecognition/.

[24] G. for Developers, *Pose landmarks detection task guide — mediapipe*, Google Developers. [Online]. Available: https://developers.google.com/mediapipe/solutions/vision/pose%5C_landmarker.

[25] G. LLC, *Cloud text-to-speech - speech synthesis*, Google Cloud. [Online]. Available: https://cloud.google.com/text-to-speech.

[26] OpenAI, *Openai api*, platform.openai.com. [Online]. Available: https://platform.openai.com/docs/api-reference/introduction.