Medi-Watch: An Automated Patient Behavior Detection System

Youssef M. Farag*
Faculty of Computer Science
October University of Modern Sciences & Arts (MSA)
Giza, Egypt
youssef.mohamed69@msa.edu.eg

Mina Samir Ghobrial
Faculty of Computer Science
October University of Modern Sciences & Arts (MSA)
Giza, Egypt
mina.samir3@msa.edu.eg

Pola Emanuel Melad
Faculty of Computer Science
October University of Modern Sciences & Arts (MSA)
Giza, Egypt
pola.emanuel@msa.edu.eg

Steven Hany Saad
Faculty of Computer Science
October University of Modern Sciences & Arts (MSA)
Giza, Egypt
steven.hany@msa.edu.eg

Ayman Atia
HCI-LAB, Faculty of Computer Science & Artificial Intelligence
Helwan University, October University of Modern Sciences & Arts (MSA)
Giza, Egypt
aezzat@msa.edu.eg

***Abstract*—Medi-Watch aims to further improve patient care and monitoring within healthcare facilities through the integration of IoT and at a lower cost. We have developed a comprehensive patient management system that utilizes automated camera monitoring to track various aspects of patient behavior including eating patterns, sleep cycles, medication adherence, fall detection, facial expression, patient's gestures requests. This could in turn be repurposed and integrated in existing health care systems. Medi-Watch integrates all detection modules in one seamless package and ensure patient privacy by adding a factor of anonymity between the nursing staff and the patient monitored by withholding live camera feeds and only updating live statuses. This also introduces a convenience factor to the nursing staff as they receive real-time notifications via the desktop app and the companion mobile app, enabling them to tend to the patients' needs in a quicker manner. The proposed system with proper integration resulted in a combined average of 92.58% accuracy in all detection categories.**

I. Introduction

In the ever-evolving landscape of healthcare technology, the objectives of Medi-Watch is to step up the standard of care in medical facilities at a low hardware cost. Our fully automated system integrates camera technology with the support of the open-source framework Media-pipe to monitor patients' behaviors seamlessly within their rooms with the exception of an LSTM-based approach for the fall detection category. This comprehensive approach encompasses vital aspects of patient well-being in multiple categories, including eating patterns, sleep cycles, medication adherence, fall detection, facial expression tracking, and responsiveness to hand gesture requests, in which we call them "Soft Behaviors" or behaviors that fall out of disease related categories. Severe burnout syndrome is an ongoing dilemma with 50% of care physicians and one-third of critical care nurses (Embriaco et al., 2007) [14]. This project could in turn help alleviate this by introducing a factor of convenience, instead of manual monitoring efforts, the system provides full automation of patient behavior monitoring. The system can also be utilized with vital sign monitors, which alone, might not be sufficient for comprehensive patient tracking as (Bickley et al., 2018) [15] has shown. Central to Medi-Watch's functionality is its full support for patient privacy and confidentiality. The system only captures crucial data points through camera surveillance per room, processing them, and deleting those frames, keeping only the log files. This helps maintain the integrity of the patient's personal space. Rather than relaying live camera feeds to nursing staff, the system only sends real-time patient statuses, ensuring that sensitive visuals remain confidential. Empowering nursing staff with timely insights and actionable information, our system also consists of an alert mechanism that seamlessly communicates with the Medi-Watch desktop and mobile application. Another aspect of Medi-Watch, As of the time of writing this paper, there is no other proposed architecture that offers this much integrated detection modules, adding to that the integration with advanced means of communication like mobile devices.

II. RELATED WORK

In this section, we reflect on the prior literature on automating the detection of patient behaviors and situate, within its canon, the agenda of building a fully automated system that is human-centered in nature with minimal costs and highly reliable. This agenda has remained a challenge due to the complexity of human behavior. We synthesize this research to suggest that AI and automation, though not sufficient by themselves, open a new angle for creating highly convenient and reliable solutions. Due to limitations in the length of this paper, we were unable to comprehensively compare each modules in different papers with what Medi-Watch integrates due to unreleased implementations and the lack of comparable datasets. We chose to highlight the parts that are more prominent like previous privacy and emergency fall detection implementations.

A. Video Object Detection for Privacy-Preserving Patient Monitoring in Intensive Care.

Although the authors didn't focus mainly on the behaviors they were detecting, initially (Emberger et al. 2023) [7] proposed the idea of blurring the ICU's camera feeds. By blurring the video footage, sensitive information such as patient identities, medical conditions, and staff interactions are obscured. The paper introduces a tweaked YOLOv5 version that works by repurposing the RGB channels of the frames and providing the useful information about the temporal consistency of the frame succession (and thus object movement). Different patient motions can be detected and medical personnel presence can be detected in the room, that can help improve the safety and quality of services for patients. The behaviors that have been identified in this research are "out of place" behaviors which means the movement of a patient within the room, like sitting, standing, or walking. It wouldn't be accurate enough to capture more details or identify multiple occurring behaviors due to the added blur. this can help us to measure and improve the level of comfort of the patient. Our approach offers a different solution, rather than blurring then processing, we process then drop the frames. Limitations faced are the blurring of the video of the object detection and the fixed position of the camera viewing, also certain types of care units that will not gather all possible scenarios will occur. In addition, authors have approved that their ways of approach have lost the color information that plays a very important role in determining object classes.

B. An Approach to Real-Time Fall Detection Based on OpenPose and LSTM

Fall Detection is an important and highly sensitive task to take on to ensure the safety of patients. In this research paper Authors find a new approach for real-time fall detection using OpenPose, which has performed state-of-art performance and using models like LSTM and RNN that store temporal dependencies, Authors performed this task using OpenPose to extract 15 key points from the human body stored in images after that simple filtering and normalizing process happen to images to enhance the quality of them. These coordinates will be given to the LSTM model that can learn the patterns for human motion and classify it into two categories which are fall and not fall. The proposed approach didn't use any wearable devices as cameras sufficed in capturing details. Although the paper doesn't provide specific numerical scores, a similar Open-Pose LSTM implementation showed a 98.2% accuracy on an undisclosed dataset 570 sequences of 30 frames each, recorded in five different rooms from eight different viewing angles by (Lin et al., 2021). [8]. We took a similar approach, but used Media-Pipes's pose landmarks to capture skeletal coordinates and feed it to the LSTM model. We have no data regarding the superiority or comparison results due to the unreleased implementation. Though in theory, it should not affect results as we used the same approach in the classification part. The limitations faced in both papers is that OpenPose required high computational power and will affect real-time implementation. Scores may also largely vary from dataset to dataset due to there not being a set standard to a diverse one. As a result the models showed signs of overfitting indicating the limitations of the training process and noise was added to alleviate it, Addressing the problems will enhance the robustness of the proposed fall detection approach (Chen et al., 2022) [8].

III. METHODOLOGY

Medi-Watch is composed of several behavior detection and interaction models that run on multiple cameras, capturing each frame and processing it. The following tests were conducted using a generic HD webcam hooked to a laptop. However, in a standard healthcare facility, multiple instances are running to accommodate the number of rooms and currently registered patients. Detection models include TUIO for the detection of medications administered, hand gestures to enable patients to make requests, sleep detection, facial expression monitoring, eating detection, and emergency fall detection.

A. TUIO Technology

TUIO (Tangible User Interface Objects) technology is used to enhance medication management for elderly people. The camera installed is utilized to detect the QR code on the medication packages as in Figure 1, Due to the ability of QR codes to pack highly encoded data (Ferano et al., 2022) [6], it was the best candidate for fast recognition and retrieval of data. The scanning process is done by using a Python library Pyzbar. Also using the same library, we can generate several QR codes as nurses specify for each patient. Once the QR code has been detected through the in-room camera, the system will notify nurses that the patient has taken the specified medicine, and the nurse can check in daily to see the medications administered by each patient Figure 1 below.

B. Hand Gesture

As gestures become more of a convenience factor in interactions (Harris et al., 2021) [5]. We have integrated Hand Gesture Technology to help nurses know if a request has been made by a patient. Cameras will detect the hand gestures performed by the patient, current gestures include: open palm and closed fist for simplicity. Once an open palm request is initiated, a notification is sent to the nearest caretakers with the patient's details and room number, a patient can cancel this request if

Fig. 1. TUIO technology.

initiated accidentally using the fist gesture. A demo is shown in Figure 2 The classification library is Google's Media-pipe. Based on the landmarks drawn on the patients hands we define if a request is made or not. We tested using (Maqueda et al., 2015) [4] proposed dataset.

1) Dataset: . The dataset is composed of two parts set 1 and set 2. In set 1 there are five hand gestures which are palm, cursor, left click, right click, and fist, and set 1 consists of the training folder and test folder both having the five hand gestures. Also, set 2 contains five hand gestures and contains a .mat file for MATLAB format. In the open palm and fist category we received a score of 93.49% accuracy.

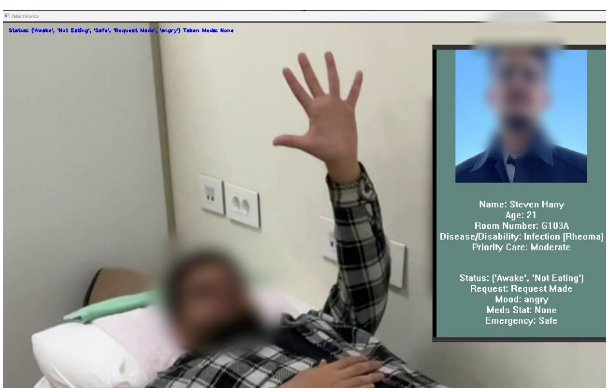

Fig. 2. Request Made by Hand Gesture

C. Fall Detection

1) Dataset: (FALL-UP) is the dataset used for fall detection. The database shows 17 participants performing 11 distinct tasks, such as five everyday activities and six different kinds of falls. The dataset is a total of 812 GB in size. Five wearable sensors were used to acquire the data: an EEG brain sensor, six pairs of infrared sensors, two cameras, and sensors positioned at the left ankle, right wrist, neck, waist, and right pocket. The cameras were positioned to record images of the falls and activity on the left and front sides of the subject, originally created by (Lourdes Martínez-Villaseñor et al., 2019) [1].

2) Data Labeling: The proposed method uses only computer vision so the only data that was needed is the frames captured from both cameras. The proposed fall detection method is made to detect if the person is flat-lined at any position on the ground or starting to fall as shown in Figure 3 on the other hand the model also detects if the person is safe as shown which means standing, sitting, or doing any other actions. Due to the enormity of the dataset, a random sample of size 4.20 GB was used to train the model, the sample contains 8977 frames that were manually classified as fall and safe, 4438 frames of total size 2.9 GB are classified as fall and 4539 of total size 2.11 GB are classified as safe, each category was placed in a separate folder. Each frame was then processed using the mediapipe pose, the result of the mediapipe pose is (x, y, z, visibility) (C. A. Q. Bugarin et al) [2] for each pose of the 33 landmarks, for each detected pose landmark which was stored in a CSV file and a label next to it 0 for fall and 1 for safe.

3) Fall detection approach: The model is constructed using several layers of long short-term memory (LSTM) which is a deep learning method, LSTM is a type of recurrent neural network (RNN) [3]. The model contains an input layer that takes an input of 33 landmarks of each pose and each landmark has 4 features which are (x, y, z, visibility) [2], followed by some LSTM layers which are followed by a dense layer with a sigmoid-activation function for binary classification. The model uses a binary cross-entropy loss function and is compiled using Adam optimizer and accuracy is chosen as the evaluation metric. Also, the data set was divided into training testing and validation at 70:15:15 respectively.

4) Fall detection Results: The Evaluation metrics used on the training data were accuracy, precision, recall, and F1 Score and achieved 0.9958, 0.9912, 1.0, 0.9956 respectively. A confusion matrix is provided as shown in Figure 4. The model was trained on the dataset for 150 epochs and achieved a training accuracy of 0.9977 and validation accuracy of 0.9958 on the data set as shown in Figure 5.

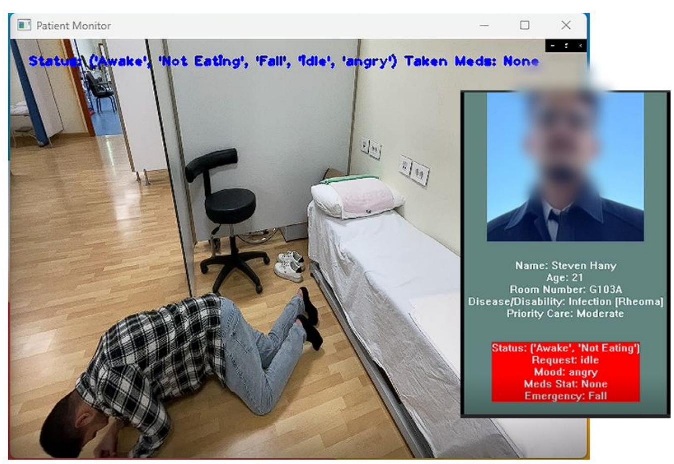

Fig. 3. Fall

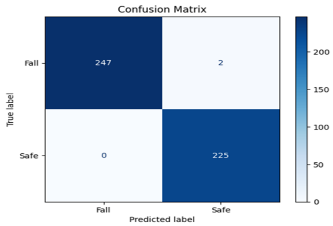

Fig. 4. Confusion matrix of the Fall detection Model

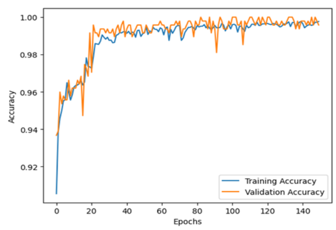

Fig. 5. Training and validation accuracy during 150 epochs

D. Sleep Detection

1) Methodology: Monitoring of sleep patterns is important in the assessment of patient health care. According to (Singh et al., 2023) [9], We can detect if an eye is open or closed using the Euclidean eye aspect ratio. The approach involves facial tracking through the face mesh of Media-Pipe as shown in Figure 6. In more detail, the main facial landmarks are detected and tracked, particularly those related to the two ocular regions. Next, for each eye, we compute the Euclidean distances between the midpoints of these landmarks in a horizontal and vertical direction. The right side and the left side thus get associated distances which are then used to form the ratios for each eye achieved by dividing the horizontal distance by the vertical distance for each eye individually. The ratios of both eyes are then averaged to obtain a final ratio. Once this final ratio is defined, when it is above a threshold (3) the patient is classified as being in a state of sleep.

2) Dataset: We tested this approach on the Closed Eyes In The Wild (CEW) dataset. According to (Tan et. al, 2014) [10], this dataset contains 2423 images split into 1192 images for people with both eyes closed and 1231 subjects with eyes open.

Fig. 6. Sleep Detection

E. Facial Expression Detection

1) Methodology: By recognizing the facial expressions of the patient this can help detect abnormal conditions in patients (Turcian et al., 2023) [11]. That's why facial expression integration was an important addition to our system. We use the Deep-Face module to recognize and analyze facial attributes and the outputs include emotion classes, each one associated with a confidence score. The emotion class with the largest confidence score will be the representative emotion of the patient.

2) Dataset: We tested this method on the FER-2013 dataset. According to (Ian et al., 2013) [12], FER-2013 is from a Kaggle competition, which is the Facial Expression Recognition Challenge 2013. This dataset consists of 35,887 48x48 pixel grayscale images of faces split into 28,709 training set and 7178 for the test set. There are 7 classes: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. We have evaluated the method on the training set.

F. Eating Detection

1) Methodology: The dietary habits of people directly impact their health conditions. So, monitoring and classifying if the patient has eaten or not is important. We use a pre-trained object detector model from MediaPipe called EfficientDet-Lite2. This model has been trained on the COCO dataset, a large-scale object detection dataset that contains 1.5 million object instances and 80 object labels. We use this model to detect any kind of food or dining utensils that exist in the frame and if anything of this is detected it will be classified as eating otherwise it is considered doing another activity. We extracted the 16 classes related to food from the 80 labels. These classes include utensils (fork, knife, spoon, bowl), various food items (banana, apple, sandwich, orange, broccoli, carrot, hot dog, pizza, donut, cake), as well as beverage containers (cup, bottle). Subsequently, we then give each frame to the model to detect objects and then check if any of the detected objects are one of the 16 food classes then we classify it as eating otherwise it will be classified as not eating.

2) Dataset: We tested this method on the Human Action Recognition Dataset from Kaggle. This dataset consists of 12,600 images split into 10,710 for the training set which we have tested on and 1890 for the test. This dataset contains 15 classes for each set. The classes are calling, clapping, cycling, dancing, drinking, eating, fighting, hugging, laughing, listening to music, running, sitting, sleeping, texting, and using a laptop.

G. Gaze Tracking

Gaze tracking is set on the Medi-Watch desktop app, in the patient's dashboard section. The webcam above the monitor checks which area of the screen a nurse is concentrating on. A heatmap is generated from the iris movements overtime to score the nurse's attentiveness on the dashboard screen. We use the face mesh of MediaPipe to get the facial landmarks as shown in sleep section in Figure 4 to take specifically the iris landmarks of both right and left eyes. Then, we average the x and y coordinates of both left and right irises and consistently record them. Using the Seaborn library, The recorded values then will be used to generate a bright-colored heatmap on the visualization canvas with higher density in those areas generally visited by the iris.

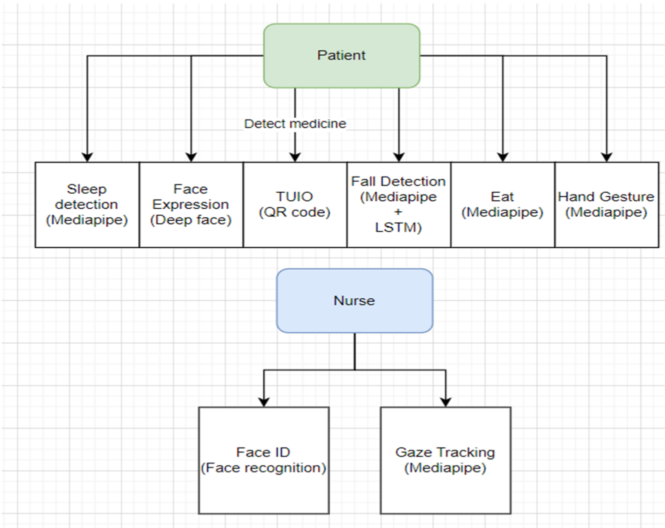

Fig. 7. Modules Accommodated

IV. Discussion

These are the scores for each detection model followed by the dataset used in the testing process. Using cameras to detect

Behavior	***Accuracy percentage***	***Dataset***
Hand Gesture	93.4965%	Gest (SET1)
Fall Detection	99.58%	FALL-UP
Sleeping	94.5934%	CEW
Facial Expression	89.3622%	FER
Eating	81.6246%	HAR

Fig. 8. Results Table

the discussed behaviors and interactions seems to be a viable option that is accurate enough and could help cut down in costs. The reason we chose this architecture with the MediaPipe framework is the low utilization and efficiency gains we get, which is a necessity for processing all the footage incoming from multiple 720p resolution cameras. Another reason is the state of the art, low latency tracking which we require for fast response. However, this didn't apply The Fall detection model, we found the LSTM approach to be more accurate and reliable. The FALL-UP dataset testing is indicative of the robust performance in the discrimination of safe poses from fall poses. In the same manner, facial tracking with MediaPipe's face mesh is applied to controlling soft behaviors like sleeping and eating. Provision of these real-time notifications and comprehensive profiles of the patient through Medi-Watch leads to optimizing care delivery. Improving ease of access and depth of information, to be in line with the changing landscape of healthcare technology, contributes to patient-centered healthcare. Future could have More improvements in every constituent of the system to stay on the cutting edge of patient care and healthcare delivery. The potential disadvantages of such systems is the high computation with minimal data transfer latency required, being able to setup a quick and robust network connected to cameras in multiple rooms is of high importance. Another disadvantage is occlusion of objects, patients may be obstructed in a way which the system may lose track and stops reporting, that's where wearable or LiDAR technologies might come in to play. In future revisions, we intend to include a more secure RSA based encryption with key exchange on the processed frames to add a higher layer of security.

V. User Interface (UI)

Medi-Watch has a desktop app and a companion mobile app. the desktop app was made with Custom TKinter (CTK) and the mobile app was made with React-Native. The database we sync across is MongoDB, which has fast read and writes and offers extreme flexibility.

1) Login: First, we start with the login page which has two options: either enter nurse ID and password or use the face ID. To identify the nurse who tries to login using face ID, we use DeepFace for face recognition. According to (Taigman et al., 2014) [16], it achieved an accuracy of 97.35% on the Labeled Faces in the Wild (LFW) dataset which contains 13,323 web images of 5,749 celebrities which are divided into 6,000 face pairs in 10 splits. On the mobile version of Medi-Watch, we used the built IR face ID which is more secure. This indeed may not be a secure form of authentication, but it was added for a proof of concept, and could be changed in future implementations to use IR-based face ID technologies as it gets more main stream.

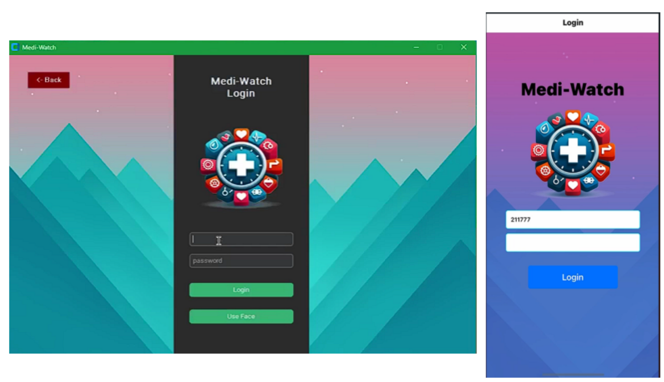

Fig. 9. Login.

2) Dashboard & Functionalities: After logging in, the patient dashboard figure 10 will be displayed as the default page which contains all patients with brief information about them and their live status. On this page, we can delete a patient, view more information about a patient, edit their information, register new patients, view active nurses, and view heatmap for nurse eye activity. However, the desktop version offers more customization and administrative access features like add, remove, and edit patients.

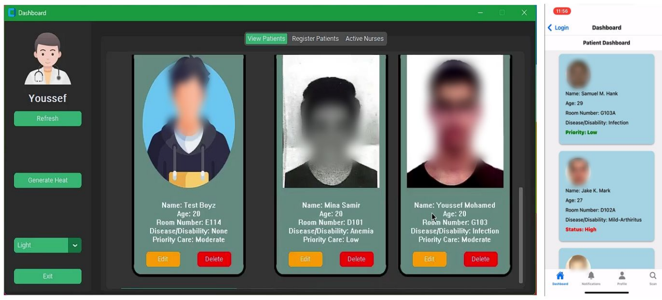

Fig. 10. Dashboard

There is also a registration section nurses could access to register patients. In the registration process, the nurses could set an initial priority care of [Low, Moderate, High] to each patient. Registration can only be done only the desktop app.

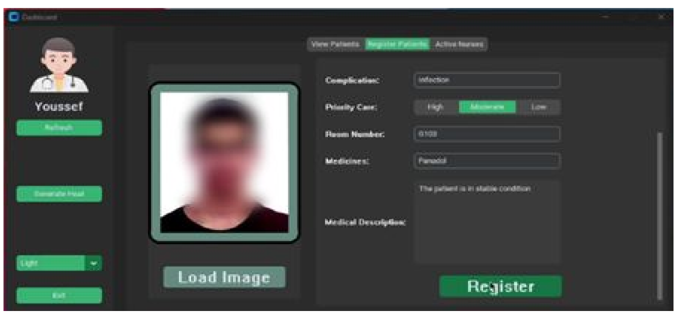

Fig. 11. Registering New Patient.

3) Active Medical Staff Page: In Figure 12 Current active medical staff can be viewed in this page with their respective details figure 12.

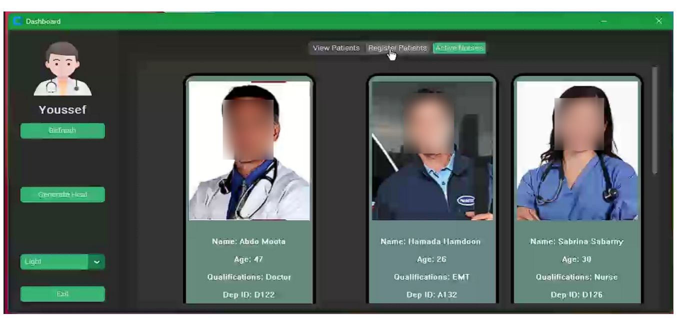

Fig. 12. Medical Stuff Page

Conclusion

To Conclude, the development of patient tracking solutions constitutes a significant advancement in the healthcare field, further research in such fields offer the potential to revolutionize patient care delivery and monitoring. Through the integration of more robust sensor technologies, advanced data analytics, and remote monitoring capabilities, these systems enable healthcare providers to obtain comprehensive insights into patient health status, behavior, and trends. In future iterations, it is crucial to prioritize patient privacy and security considerations in the design and implementation of patient tracking systems.

References

[1] Lourdes Martínez-Villaseñor, Hiram Ponce, Jorge Brieva, Ernesto Moya-Albor, José Núñez-Martínez, Carlos Peñafort-Asturiano, “UP-Fall Detection Dataset: A Multimodal Approach”, Sensors 19(9), 1988: 2019, doi:10.3390/s19091988.

[2] Bugarin, C. A. Q., Lopez, J. M. M., Pineda, S. G. M., Sambrano, M. F. C., & Loresco, P. J. M. (2022, September). Machine vision-based fall detection system using mediapipe pose with iot monitoring and alarm. In 2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC) (pp. 269-274). IEEE.

[3] Chen, P. C., Chang, C. H., Chan, Y. W., Tasi, Y. T., & Chu, W. C. (2022, June). An Approach to Real-Time Fall Detection based on OpenPose and LSTM. In 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC) (pp. 1573-1578). IEEE.

[4] A.I. Maqueda, C.R. del Blanco, F. Jaureguizar, N. García, “Human–computer interaction based on visual hand-gesture recognition using volumetric spatiograms of local binary patterns”, Computer Vision and Image Understanding, Special Issue on Pose & Gesture, vol. 141, pp. 126-137, Dec. 2015. (doi:10.1016/j.cviu.2015.07.009)

[5] Indriani, Harris, M., & Agoes, A. S. (2021). Applying Hand Gesture Recognition for User Guide Application Using MediaPipe. Proceedings of the 2nd International Seminar of Science and Applied Technology (ISSAT 2021). https://doi.org/10.2991/aer.k.211106.017

[6] Francisco Calvin Arnel Ferano , Jan Keane Olajuwon , Gede Putra Kusma (2022). QR Code Detection and Rectification Using Pyzbar and Perspective Transformation, 15th November 2022. Vol.100. No 21

[7] Emberger, Raphael, et al. “Video Object Detection for Privacy-Preserving Patient Monitoring in Intensive Care.” ArXiv (Cornell University), 2023, https://doi.org/10.48550/arXiv.2306.14620

[8] P. -C. Chen, C. -H. Chang, Y. -W. Chan, Y. -T. Tasi and W. C. Chu, ”An Approach to Real-Time Fall Detection based on OpenPose and LSTM,” 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC), Los Alamitos, CA, USA, 2022, pp. 1573-1578, doi: 10.1109/COMPSAC54236.2022.00250. keywords: Training;Time series analysis;Sociology;Neural networks;Medical services;Software;Real-time systems;OpenPose;Deep learning;Fall Detection;Image recognition,

[9] Chuan-Bi Lin, Ziqian Dong, Wei-Kai Kuan, Yung-Fa Huang, ”A Framework for Fall Detection Based on OpenPose Skeleton and LSTM/GRU Models”, Appl. Sci. 2021, 11(1), 329; https://doi.org/10.3390/app11010329

[10] Singh, J., Kanojia, R., Singh, R., Bansal, R., & Bansal, S. (2023, March 11). Driver Drowsiness Detection System: An Approach By Machine Learning Application. ArXiv.org. https://doi.org/10.48550/arXiv.2303.06310

[11] F.Song, X.Tan, X.Liu and S.Chen, Eyes Closeness Detection from Still Images with Multi-scale Histograms of Principal Oriented Gradients, Pattern Recognition, 2014.

[12] Turcian D, Stoicu-Tivadar V. Real-Time Detection of Emotions Based on Facial Expression for Mental Health. Stud Health Technol Inform. 2023 Oct 20;309:272-276. doi: 10.3233/SHTI230795. PMID: 37869856.

[13] Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.-H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Shawe-Taylor, J., Milakov, M., Park, J., & Ionescu, R. (2013). Challenges in Representation Learning: A report on three machine learning contests. ArXiv:1307.0414 [Cs, Stat]. https://arxiv.org/abs/1307.0414

[14] N. Embriaco et al., ”High level of burnout in intensivists: Prevalence and associated factors, Am. J. Respir, vol. 175, no. 7, pp. 686-692, 2007.

[15] Brekke IJ, Puntervoll LH, Pedersen PB, Kellett J, Brabrand M (2019) The value of vital sign trends in predicting and monitoring clinical deterioration: A systematic review. PLoS ONE 14 (1): e0210875.https://doi.org/10.1371/journal.pone.0210875

[16] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, ”DeepFace: Closing the Gap to Human-Level Performance in Face Verification,” 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp. 1701-1708, doi: 10.1109/CVPR.2014.220.