Computer Vision Based Approach for Patient Gait Abnormality Detection and Monitoring

Bandaru Jaya Nandini, Mamidi Leha Sahithi, Siddareddy Gari Harshika, Radha D, Nandu C Nair
Department of Computer Science and Engineering
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
bl.en.u4aie22006@bl.students.amrita.edu, bl.en.u4aie22035@bl.students.amrita.edu, bl.en.u4aie22053@bl.students.amrita.edu,
d_radha@blr.amrita.edu, c_nandu@blr.amrita.edu

***Abstract*—Irregularities of gait are an indicator of a range of clinical conditions such as Parkinson's disease, stroke, cerebral palsy, and orthopaedic conditions. Early recognition of such patterns is vital for prompt diagnosis, rehabilitation planning Current gait analysis depends on costly hardware or human observation, making it unscalable and inaccessible. We address this with an automated computer vision-based solution that inspects walking videos to identify abnormal gait patterns. The system employs the Gait Abnormality in Video Dataset (GAVD), a large-scale clinically annotated video dataset, as the basis for model training and testing. Our pipeline starts with human detection with YOLOv8, followed by pose keypoint extraction with MediaPipe. These keypoints are ordered to record temporal gait dynamics, which are further classified with the help of a Transformer encoder model into eight types such as normal, stroke, Parkinson's, cerebral palsy, prosthetic, myopathic, and exercise-related gait. The overall solution is implemented as a Streamlit web app, with support for real-time video input and instant predictions. For clinically relevant abnormal patterns, an alert is sent through WhatsApp using the Twilio API, ensuring immediate medical care. This strategy provides an affordable, scalable, and non-invasive way to perform automatic gait analysis on real-world video data. The experimental results achieved high classification performance, and the Transformer model produced a test accuracy of 95.72%, which surpassed the other models in precision, recall, F1-score, and balanced accuracy.**

***Index Terms*—Gait Abnormality, Computer Vision, Human Pose Estimation, MediaPipe, Transformer Encoder, GAVD Dataset, Streamlit Web App.**

I. INTRODUCTION

Gait abnormality is normally indicative of associated neurological or musculoskeletal disease conditions like Parkinson's disease, stroke, cerebral palsy, and orthopedic disorders. Recognition of such abnormality in the initial phase is of extreme importance in clinical assessment, rehabilitation planning, and also prevention of prospective injury. Traditional gait analysis traditionally depends upon motion-capture systems, force platforms, or visual inspection, all of which necessitate costly equipment, highly qualified personnel, and an otherwise controlled laboratory setup. These constraints make the use of gait monitoring in remote or field settings inappropriate, especially for instances of extended or continuous patient monitoring.

To address this, automated video-based gait analysis using computer vision offers a scalable, non-invasive alternative. The solution pipeline utilizes the Gait Abnormality in Video Dataset (GAVD), a publicly available dataset composed of over 450 YouTube videos containing annotated gait categories verified by clinical experts. This dataset includes more than 416,000 labeled frames and over 458,000 bounding box coordinates across gait classes: Normal, Abnormal, Stroke, Parkinson's, Cerebral Palsy, Myopathic, Prosthetic, and Exercise-induced gaits. To ensure class balance and efficient model learning, a stratified subset of the dataset selected, with approximately 10,426 frames retained for each class to create a balanced and reduced training set.

The pipeline begins by applying YOLOv8 for accurate person detection across frames. Detected subjects are then passed through MediaPipe Pose to extract 99-dimensional 3D keypoint vectors. These frame-wise pose landmarks are temporally combined into fixed-length sequences, capturing gait dynamics over time. A Transformer Encoder model** processes these pose sequences to classify the subject into one of the 8 clinically defined gait categories. The system is designed to handle real-world walking videos, including those with varying camera angles and backgrounds.

To evaluate the performance of the Transformer-based model, comparative experiments were conducted using two additional architectures: CNN + BiLSTM and TCN + BiLSTM. Results showed that the Transformer model significantly outperformed the others, achieving a test accuracy of 95.72%, with high precision, recall, F1-score, and other evaluation metrics. This classification system is fully deployed as a Streamlit web app, which enables video uploads, gait prediction, and automatic WhatsApp alerts (via Twilio API) for abnormal cases. The integration of stratified dataset preprocessing, keypoint-based sequence modeling, and deep learning classification provides an end-to-end solution for accessible and clinically relevant gait abnormality detection.

This system not only facilitates real-time gait monitoring but also closes the gap between clinical gait assessment and realistic remote deployment. By using light models such as YOLOv8 for detection, fast keypoint extraction using MediaPipe, and the strong sequence modeling power of the Transformer architecture, the system maintains accuracy as well as computational affordability. The architecture allows for early detection of gait-related health issues with minimal infrastructure cost, making it an effective tool for scalable healthcare applications in low-resource and telemedicine environments.

II. Literature Survey

Chen et al. [1] investigated the use of machine learning models—SVM, KNN, CNN, and LSTM—for classifying flatground gait patterns captured using Kinect-based motion systems. Their dataset consisted of 750 samples representing simulated gaits (normal, pelvic obliquity, and knee hyperextension), and the SVM model achieved the highest accuracy of 94.9Multi-Level Fine-Tuned Transformer model on the CASIAB dataset for identity-based gait recognition, demonstrating the effectiveness of multi-layer attention mechanisms in capturing spatiotemporal gait features across different conditions and view angles. Ranjan et al. [5] introduced the GAVD dataset—comprising over 450 clinician-annotated videos from YouTube—and benchmarked its utility using TSN and SlowFast networks, achieving up to 94 Several studies have explored the application of pose estimation techniques for markerless gait analysis. Tony Hii et al. [3] utilized OpenPose and BlazePose to extract 2D joint keypoints from frontal-view gait videos, enabling the calculation of joint trajectories, stride length, and step width variability. A subsequent comparative study by the same authors [7] evaluated OpenPose, MediaPipe, and MMPose for markerless joint tracking, with MediaPipe demonstrating the best balance between accuracy and real-time performance.

Vats et al. [13] focused on children with cerebral palsy and developed a sagittal-plane gait analysis method using pose estimation to assess hip, knee, and ankle movements without the need for wearable markers. Clinical validation of video-based gait analysis systems has also been a focal point. Lonini et al. [12] conducted a proof-of-concept study involving stroke survivors, where gait videos were analyzed using OpenPose and validated against the Vicon motion capture system. Their results showed a 3.5 cm mean error in step length estimation and a 93in gait cycle timing. Talaa et al. [8] designed a home-based rehabilitation assessment framework using RGB videos and computed Dynamic Time Warping (DTW) scores to monitor recovery progress, achieving clear differentiation between normal, abnormal, and recovering gait profiles. Both studies highlight the clinical potential of markerless systems while noting challenges such as lighting conditions, occlusions, and the absence of real-time feedback. Pose-based activity recognition has been taken beyond gait into other organized human motion spaces. Srivastava et al. [6] applied a hybrid CNN + LSTM model with MediaPipeextracted keypoints for detecting yoga poses with 9510,000 annotated yoga pose images, emphasizing diversity in body types, environments, and angles. These works underline the adaptability of pose estimation frameworks across various motion classification tasks and demonstrate the robustness of keypoint-based models when supported by well-annotated datasets. Temporal gait and motion sequence classification has been tackled via deep learning model comparisons.

Lee et al. [11] compared 1D-CNN, GRU, and LSTM architectures for the classification of step duration in adolescent and older groups. Their results showed that 1D-CNN achieved the highest accuracy, while GRU offered efficient temporal learning with reduced computational demand compared to LSTM. D ́ıazArancibia et al. [9] presented a scalable gait analysis system combining OpenPose, YOLOv3, and ResNet50, deployed via Docker and AWS, and validated using Kinovea software. Their framework featured a GUI for clinical specialists and emphasized remote accessibility and scalability. Broader advancements in gait recognition also include enhancements to dataset representation and model generalization. Kour et al. [10] provided a survey of computer vision-based Parkinson's Disease gait recognition methods, highlighting the use of spatiotemporal and kinematic features along with hybrid CNN–RNN models. Bastos and Tavares [14] incorporated joint angle data into gait sequence modeling on CASIAB and CASIA-E datasets using their GaitBase model variants, improving recognition accuracy under different settings.

Agarwal et al. [4] applied the YOLOv8 object detection model to automate fracture detection in radiographic images, showcasing the utility of deep object detectors in clinical applications—a concept mirrored in the use of YOLO-based models for detecting humans in gait video preprocessing pipelines.

III. Methodology

The proposed system follows a structured computer vision pipeline to detect and classify gait abnormalities from walking video input. The methodology is divided into several key stages, each playing a vital role in achieving accurate and clinically relevant gait classification.

A. GAVD Dataset Acquisition

The system is built on the foundation of the publicly accessible Gait Abnormality in Video Dataset (GAVD), which consists of more than 450 clinically labeled gait videos. These videos are obtained from online resources and cover varying conditions such as Parkinson's disease, stroke, cerebral palsy, myopathy, prosthetic gait, and others. Each video is linked to frame-level annotations and video IDs, while the dataset contains over 416,000 labeled frames. In order to provide balanced representation during training of the model, a stratified subset of around 10,426 frames for each class was chosen, for a total of 79893 frames. Through this preprocessing task, effective learning is achieved along with clinical diversity across gait classes.

The Gait Abnormality in Video Dataset (GAVD) as shown in Fig 1 consists of eight clinically labeled gait classes derived from more than 450 authentic walking videos. The categories are normal, abnormal, cerebral palsy, stroke, Parkinson's, myopathic, prosthetic, and exercise-induced gait. These categories are varied and cover a variety of neuromuscular and biomechanical impairments as well as normal movement, allowing for thorough training and testing of gait classification models. The dataset is frame-level annotated and spans diverse

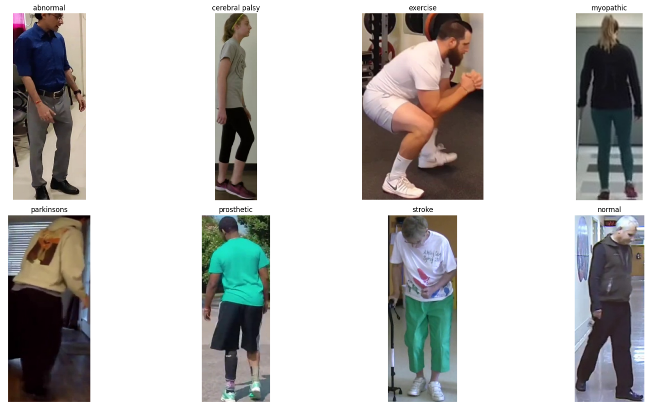

Fig. 1. Sample frames of eight gait categories in the GAVD

walking conditions, adding to real-world relevance and strong model generalization.

B. Human Detection and Frame Processing

From the provided YouTube video URLs, frames are extracted using OpenCV. To isolate the walking subject, YOLOv8 - a state-of-the-art object detection model is applied to each frame. Only regions corresponding to the person class are retained. This ensures that only valid walking individuals are processed, eliminating background noise and irrelevant content.

C. Pose Keypoint Extraction

The MediaPipe Pose model is then used to extract 33 3D anatomical landmarks (x, y, z) from each detected human frame, resulting in 99-dimensional pose vectors per frame. Normalization is applied to handle scale and viewpoint variation. If the system detects the person is facing backward (via shoulder and hip orientation), a correction is applied by swapping corresponding left and right joints.

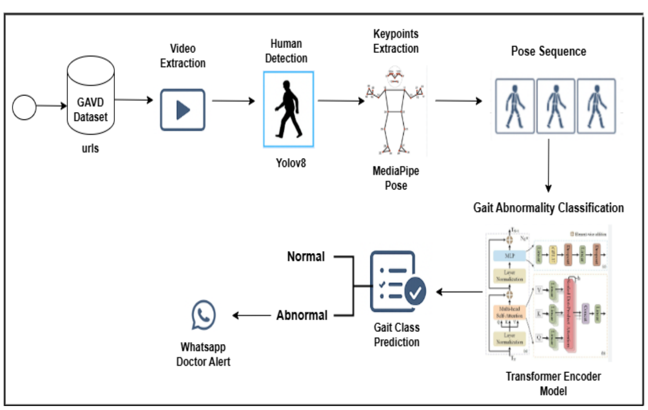

Fig. 2. System architecture for gait abnormality detection

D. Pose Sequence Generation

Since gait is a temporal process, isolated frames are insufficient. To capture motion dynamics, consecutive pose keypoints are grouped into fixed-length sequences (typically 5 frames). Each sequence encodes spatial and temporal gait features, forming the input to the classification model. Sequences are generated using a sliding window method to ensure overlap and smooth transitions.

E. Gait Classification with Transformer Encoder

A Transformer Encoder model is trained to classify the gait sequence into one of the eight categories. The transformer architecture, which includes multi-head self-attention and positional encoding, captures long-range dependencies between frames. The input is a sequence of 99-dimensional vectors, and the output is a softmax distribution over the gait classes.

F. Prediction and Doctor Alert

Upon classification, if the predicted label falls under abnormal categories (e.g., Stroke, Parkinson's, Myopathic), a WhatsApp alert is triggered using the Twilio API. Each abnormal class is mapped to a corresponding specialist (neurologist, orthopedist, etc.) based on a predefined routing strategy. This mechanism facilitates remote patient monitoring and timely intervention.

G. Streamlit Web-Based Deployment

The complete solution is integrated into a real-time Streamlit web application. Users can upload any walking video, and the app performs detection, classification, and alerting in an end-to-end manner. The interface provides visual feedback of predictions, making it accessible for both clinical and non-clinical use cases.

H. Model Evaluation

The Transformer Encoder model is tested on a held-out test set with standard classification metrics: Accuracy, Precision, Recall, F1-Score, Balanced Accuracy, Matthews Correlation Coefficient (MCC), and Cohen's Kappa Score. These metrics give a complete evaluation of the performance of the model, particularly in multiclass clinical scenarios where class imbalance can occur. The results show excellent generalization ability and high reliability in identifying varied gait abnormalities.

I. Comparison with Other Models

To validate the effectiveness of the Transformer Encoder, its performance is compared against two other baseline architectures:

- **CNN + BiLSTM:** Combines convolutional spatial feature extraction with temporal modeling using Bidirectional LSTM layers.
- **TCN + BiLSTM:** Uses Temporal Convolutional Networks followed by BiLSTM layers for sequential modeling.

All models were tested and trained on the same stratified and balanced dataset. The Transformer model far surpassed the others in all measures, having a test accuracy of 95.72%, macro-averaged precision, recall, and F1-score of 0.96.

IV. RESULTS AND DISCUSSION

The effectiveness of the proposed gait classification system was empirically validated through rigorous experimentation involving multiple deep learning architectures. Each model was evaluated on its ability to accurately interpret temporal pose sequences derived from real-world clinical gait data. By maintaining consistency in dataset distribution and evaluation criteria, the study ensures a fair comparison across models. Emphasis was placed not only on achieving high accuracy but also on evaluating each model's performance across precision-oriented metrics crucial for clinical reliability. The comparative outcomes highlight the architectural strengths and limitations in handling complex spatiotemporal gait patterns, particularly for multi-class abnormal gait detection.

The accuracy of the gait abnormality detection system was tested on three varying model architectures: CNN + BiLSTM, TCN + BiLSTM, and the Transformer Encoder model proposed in this work. All models were trained and tested on a stratified and balanced subset of the GAVD dataset with an equal number of representations for every gait class to provide fairness.

The relative performance of the models is displayed in 2, presenting the test accuracy of each model on the classification task. The Transformer Encoder model performed much better than the baselines, with a test accuracy of 95.72%, while that of TCN + BiLSTM was 87.02% and that of CNN + BiLSTM was 67.71%.

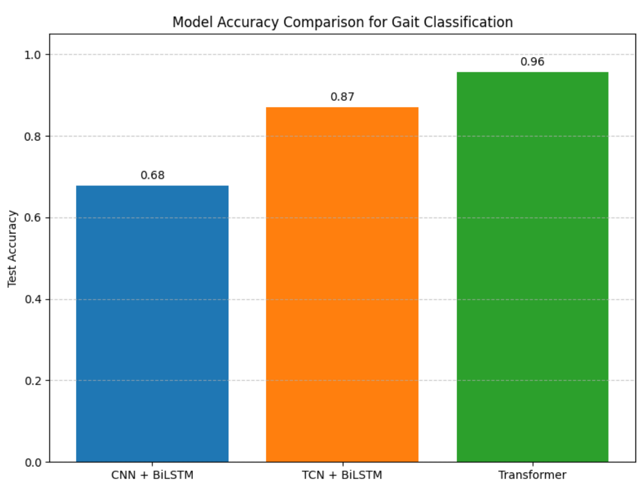

Fig. 3. Accuracy Comparision for Gait Classification

As shown in Fig 4 illustrates the custom-built Gait Detection and Doctor Alert System developed using Streamlit. The web interface allows users to upload a walking video in formats such as .mp4, .avi, or .mov. Once uploaded, the system processes the video to detect the gait pattern using pose estimation and deep learning models. If an abnormal gait (e.g., Parkinson's, stroke, or myopathic) is detected, an automated WhatsApp alert is triggered to notify the corresponding medical specialist. The app provides real-time playback of the uploaded video along with clear status feedback, making it accessible and practical for both clinical use and remote monitoring scenarios.

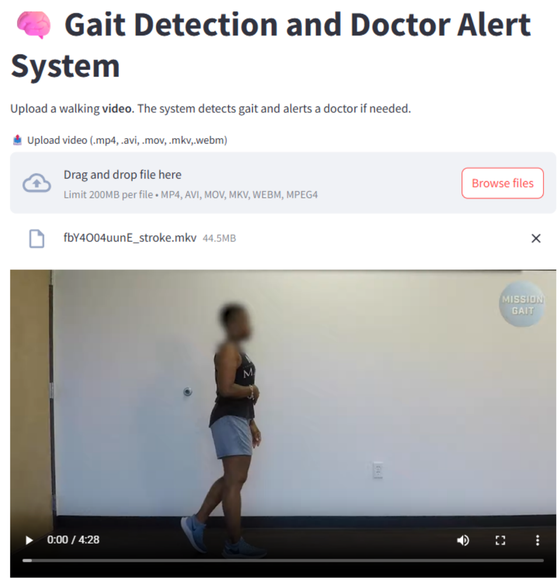

Fig. 4. web application of Gait Detection and Doctor Alert System.

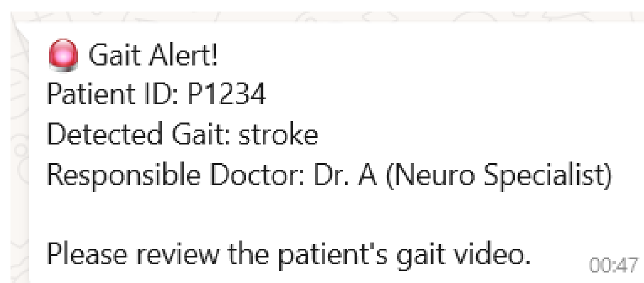

Fig. 5. Sample WhatsApp alert sent to the specialized Doctor

The system incorporates an automated doctor alert mechanism that uses the Twilio API to send real-time WhatsApp messages to medical specialists based on the detected gait abnormality. As shown in Figure 5, once a gait type is classified, a message is dispatched to the appropriate doctor: Dr. A (Neuro Specialist) handles neurological conditions such as stroke, Parkinson's, and cerebral palsy; Dr. B (Ortho Specialist) is notified for prosthetic and myopathic gait patterns; and Dr. C (General Physician) receives alerts for general abnormal gait cases. Each message includes the patient ID, predicted gait class, and a request for the doctor to review the gait video, ensuring timely intervention and effective clinical coordination.

The detailed evaluation metrics—Precision, Recall, F1-Score, Balanced Accuracy, Matthews Correlation Coefficient (MCC), and Cohen's Kappa—are summarized in Table I. The Transformer model consistently delivered superior results across all metrics, indicating not only high accuracy but also strong generalization across diverse gait types.

TABLE I
MODEL EVALUATION METRICS COMPARISON

Model	Acc.	Prec.	Recall	F1	Bal. Acc.	MCC	Kappa
CNN + BiLSTM	0.6771	0.67	0.68	0.67	0.6766	0.6309	0.6305
TCN + BiLSTM	0.8702	0.87	0.87	0.87	0.8694	0.8516	0.8513
Transformer-Encoder	**0.9572**	**0.96**	**0.96**	**0.96**	**0.9568**	**0.9510**	**0.9510**

These results validate the effectiveness of the Transformer model in capturing temporal dependencies and spatial features from the pose sequences, making it suitable for clinical applications.

V. CONCLUSION AND FUTURE SCOPE

This study presents a computer vision-based framework for detecting patient gait abnormalities using real-world walking videos. By leveraging the Gait Abnormality in Video Dataset (GAVD), human subjects were accurately detected using the YOLOv8 object detector, followed by pose keypoint extraction via MediaPipe to model body dynamics. The extracted 3D pose landmarks were organized into temporal sequences and classified using a Transformer Encoder model into clinically relevant gait categories such as Normal, Stroke, Parkinson's, Cerebral Palsy, Myopathic, Prosthetic, and Exercise-induced gait. Comparative experiments with CNN + BiLSTM and TCN + BiLSTM models demonstrated the superior performance of the Transformer-based architecture, achieving a test accuracy of 95.72% along with high values across all major evaluation metrics. The entire pipeline is deployed as an interactive Streamlit web application capable of real-time video analysis and automated WhatsApp alerts for abnormal gait detection. This end-to-end system offers a low-cost, scalable, and non-invasive tool that can be integrated into telemedicine platforms for remote patient monitoring. In future, this approach can be enhanced by incorporating multi-view pose estimation, longitudinal gait monitoring, and integration with wearable sensor data for improved robustness and diagnostic accuracy. Additionally, expanding the system to support real-time video streaming and multi-person gait analysis can significantly broaden its applicability in smart healthcare environments.

REFERENCES

[1] Chen B, Chen C, Hu J, Sayeed Z, Qi J, Darwiche HF, Little BE, Lou S, Darwish M, Foote C, Palacio-Lascano C. Computer Vision and Machine Learning-Based Gait Pattern Recognition for Flat Fall Prediction. Sensors (Basel). 2022 Oct 19;22(20):7960. doi: 10.3390/s22207960. PMID: 36298311; PMCID: PMC9612353.

[2] H. Wu and A. Zhao, ”Multi-Level Fine-Tuned Transformer for Gait Recognition,” 2022 International Conference on Virtual Reality, Human-Computer Interaction and Artificial Intelligence (VRHCIAI), Changsha, China, 2022, pp. 83-89, doi: 10.1109/VRHCIAI57205.2022.00021

[3] C. S. Tony Hii et al., ”Frontal Plane Gait Analysis using Pose Estimation Models,” 2023 IEEE 2nd National Biomedical Engineering Conference (NBEC), Melaka, Malaysia, 2023, pp. 1-6, doi: 10.1109/NBEC58134.2023.10352623.

[4] P. Agarwal and P. Kumar, ”Automated Bone Fracture Detection using YOLOv8,” 2025 3rd International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 2025, pp. 1256-1260, doi: 10.1109/ICDT63985.2025.10986404.

[5]] R. Ranjan, D. Ahmedt-Aristizabal, M. Ali Armin and J. Kim, ”Computer Vision for Clinical Gait Analysis: A Gait Abnormality Video Dataset,” in IEEE Access, vol. 13, pp. 45321-45339, 2025, doi: 10.1109/ACCESS.2025.3545787.

[6] A. Srivastava et al., ”Yoga Pose Detection with Deep Learning and Computer Vision,” 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-9, doi: 10.1109/ICCCNT61001.2024.10724343.

[7] Tony Hii, Chang Soon Gan, Kok Beng Zainal, Nasharuddin Mohamed Ibrahim, Norlinah Rani, Shahrul Abd Shattar, Normurniyati. (2022). Marker Free Gait Analysis using Pose Estimation Model. 109-113. 10.1109/SCOReD57082.2022.9974096.

[8] Talaa, Safae El Fezazi, Mohamed Jilbab, A. Alaoui, My. (2023). Computer Vision-Based Approach for Automated Monitoring and Assessment of Gait Rehabilitation at Home. International Journal of Online and Biomedical Engineering (iJOE). 19. 139-157. 10.3991/ijoe.v19i18.43943

[9] D´ıaz-Arancibia, J., Cordova, M., Arango-L ´ opez, J. et al. Automatic ´ gait analysis through computer vision: a pilot study. Neural Comput Applic (2023). https://doi.org/10.1007/s00521-023-08549-2

[10]] N. Kour, Sunanda and S. Arora, ”Computer-Vision Based Diagnosis of Parkinson's Disease via Gait: A Survey,” in IEEE Access, vol. 7, pp. 156620-156645, 2019, doi: 10.1109/ACCESS.2019.2949744.

[11]] Lee, Teng, Ezreen Farina Shair, Abdul Rahim Abdullah, Kazi Ashikur Rahman, Nursabillilah Mohd Ali, Nur Zawani Saharuddin, Nurhazimah Nazmi. ”Comparative Analysis of 1D – CNN, GRU, and LSTM for Classifying Step Duration in Elderly and Adolescents Using Computer Vision.” International Journal of Robotics and Control Systems [Online], 5.1 (2025): 426-439. Web. 17 Mar. 2025

[12] Lonini L, Moon Y, Embry K, Cotton RJ, McKenzie K, Jenz S, Jayaraman A. Video-Based Pose Estimation for Gait Analysis in Stroke Survivors during Clinical Assessments: A Proof-of-Concept Study. Digit Biomark. 2022 Jan 13;6(1):9-18. doi: 10.1159/000520732. PMID: 35224426; PMCID: PMC8832219.

[13] V. K. Vats, C. Prakash, A. Kaur and R. Kumar, ”Markerless Sagittal Gait Analysis of Cerebral Palsy in Children Using Pose Estimation Techniques,” 2025 International Conference on Innovation in Computing and Engineering (ICE), Greater Noida, India, 2025, pp. 1-6, doi: 10.1109/ICE63309.2025.10984338.

[14]] Diogo R.M. Bastos, Joao Manuel R.S. Tavares,A scalable gait ac- ˜ quisition and recognition system with angle-enhanced models,Expert Systems with Applications,Volume 269,2025,126499,ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2025.126499.

[15] M. K. HS, S. G. Ramkumar, S. N. Palavalli and S. S S, ”Yoga-Pose8 : A Diverse Yoga Pose Dataset for Classification and Analysis,” 2025 International Conference on Artificial Intelligence and Data Engineering (AIDE), Nitte, India, 2025, pp. 339-345, doi: 10.1109/AIDE64228.2025.10986902.

[16] N. S. Dhanala and R. D., ”Implementation and Testing of a Blockchain based Recruitment Management System,” 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020, pp. 583-588, doi: 10.1109/ICCES48766.2020.9138093.

[17] S. Vishwakarma, D. Radha and J. Amudha, ”Effectual Training for Object Detection Using Eye Tracking Data Set,” 2018 International Conference on Inventive Research in Computing Ap-

plications (ICIRCA), Coimbatore, India, 2018, pp. 225-230, doi: 10.1109/ICIRCA.2018.8597275.
[18] K. Das and R. D, "Implementation and Analysis of Image Compression Using Ramanujan's Sum," 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2022, pp. 1-6, doi: 10.1109/ICCCNT54827.2022.9984427.
[19] B. V. S. N. S. Vinay, R. D. and V. S. K. Devi, "Credit Card Fraud Detection(CFD) Using GNN and Ensemble Techniques," 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2025, pp. 1991-1996, doi: 10.1109/IDCIOT64235.2025.10915111.
[20] S. A. S, R. D and V. S. K. Devi, "Forecasting Bitcoin Price Trends: Integrated Machine Learning with Market Trends," 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2025, pp. 1492-1497, doi: 10.1109/IDCIOT64235.2025.10914850.
[21] A. Bhardwaj, R. D. and V. S. K. Devi, "An Enhanced Framework for Churn Prediction using Stratifed Bagging and Stacked Multi-Output Random Forests," 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2025, pp. 119-123, doi: 10.1109/IDCIOT64235.2025.10914868.