

# StrideGuard: Deep Learning-Based Detection of Gait Abnormalities for Healthcare Monitoring

Shailaja K.P<sup>1</sup>, Shilpa S<sup>2</sup>

<sup>1</sup>Assistant Professor, Department of Computer Applications, B.M.S. College of Engineering, Bull Temple Road, Bengaluru-19. Email: [shailaja.mca@bmsce.ac.in](mailto:shailaja.mca@bmsce.ac.in)

<sup>2</sup>Assistant Professor, Department of Computer Applications, B.M.S. College of Engineering, Bull Temple Road, Bengaluru-19. Email: [shilpa.mca@bmsce.ac.in](mailto:shilpa.mca@bmsce.ac.in)

ORCID: 0000000163048296 (Shailaja K.P) | ORCID: 0009000767457178 (Shilpa S)

## ABSTRACT

Video-based, contactless gait analysis provides a user-friendly alternative to wearable sensors for identifying walking abnormalities. Our method begins by segmenting RGB video into individual gait cycles and converting them into Gait Energy Images (GEIs), which encapsulate a cycle's silhouette dynamics. These GEIs are processed by a tailored Convolutional Neural Network that learns hierarchical spatial-temporal features end-to-end to classify five gait patterns: normal, antalgic, spastic, scissor, and Parkinsonian. To demonstrate effectiveness, we benchmark our CNN against Support Vector Machines and Random Forests trained on the same GEI inputs, achieving over 94% accuracy—more than a 10% gain compared to classical classifiers. We further evaluate robustness under varying covariates (clothing, camera angle) on the CASIA-B dataset and in-house clinical recordings, sustaining high sensitivity and specificity. Our lightweight architecture runs at over 40 fps on a single GPU, making it suitable for real-time applications. These results validate our AI-driven, noninvasive framework as a scalable solution for elder-care screening, rehabilitation monitoring, and surveillance in diverse, real-world settings.

**Keywords:** Gait Anomaly Prediction • Gait Energy Image (GEI) • Convolutional Neural Network (CNN) • Appearance-Based Gait Analysis • Discriminative Feature Learning • Covariate Invariance • Non-contact Monitoring • Spatiotemporal Feature Extraction.

**How to cite this article:** Shailaja KP, Shilpa S. StrideGuard: Deep Learning-Based Detection of Gait Abnormalities for Healthcare Monitoring. *Int J Drug Deliv Technol.* 2026;16(56s): 1402-1410. DOI: 10.25258/ijddt.16.56s.153

**Source of support:** Nil.

**Conflict of interest:** None.

**Spatiotemporal Feature Extraction**

## INTRODUCTION

Gait—the way in which a person walks—carries rich information about identity, health status, and functional mobility. Minor deviations in stride length, timing, or posture can serve as early indicators of musculoskeletal disorders, fall risk, or neurodegenerative decline. Traditional clinical gait assessments, however, rely on laboratory-grade equipment or time-consuming expert observation, limiting scalability and real-world applicability.

In this project, we focus on a purely video-based pipeline leveraging Convolutional Neural Networks to classify four key gait anomalies—antalgic, spastic, scissor, and Parkinsonian gaits. We record gait sequences via standard RGB cameras, automatically extract spatial-temporal features through a CNN, and then classify each walking pattern without manual feature engineering (unlike SVM or Random Forest approaches [4], [6]). Finally, we evaluate our model's predictive accuracy against benchmark datasets to demonstrate its practical utility in medical screening and rehabilitation contexts.

To address remaining limitations, we further propose a hybrid, camera-based framework that combines

appearance-based Gait Energy Images with discriminative feature learning to maintain high accuracy across both known and unseen conditions. By integrating convolutional networks for silhouette-based anomaly detection with a region-focused invariant feature extractor, our system aims to deliver robust, non-invasive gait analysis suitable for clinical screening, elder-care monitoring, and public-safety applications.

A variety of sensing modalities and analysis techniques have been explored to automate gait monitoring. Inertial wearable devices mounted on the head [1] or body [4] capture acceleration and angular velocities to flag unstable strides, while instrumented insoles record pressure and motion for feature-based classification [6]. Specialized rigs targeting trans-tibial amputees employ force and angle sensors to detect asymmetry in real time [10]. Marker less depth cameras combined with CNN + LSTM architectures offer non-contact posture recognition [5], and contrastive learning on floor-vibration signals achieves user-invariant anomaly detection without wearables [7].

Vision-based approaches likewise have advanced rapidly. The SAIL system processes Timed Up-and-Go videos with convolutional backbones and attention modules to predict clinical mobility scores automatically [2]. Monocular RGB sequences fed into SimpleHRNet plus

LSTM modules reconstruct 3D gait cycles and flag anomalies across varied viewpoints and attire [8]. Off-the-shelf 3D pose estimators further enable spatiotemporal parameter extraction—such as stride duration and joint variance—to screen for early dementia signs in elder cohorts [3]. Unsupervised clustering on wearable-sensor time series has even been used to segment gait into stance and swing phases without labelled data, closely matching supervised benchmarks [9].

## 1. RELATED WORK

Research on automated gait anomaly detection has undergone a notable transition over the last decade, evolving from heavily instrumented wearable systems to contactless, vision-based approaches and, more recently, to infrastructure-assisted and unsupervised learning solutions. Each paradigm offers distinct advantages and faces unique limitations, with recent advances focusing on improving clinical applicability, reducing user burden, and enabling deployment in naturalistic environments.

Early wearable-based solutions relied on body-mounted sensors to capture kinematic and kinetic data during walking. For example, Lin and Lin [1] demonstrated that Inertial Measurement Units (IMUs), when combined with Temporal Convolutional Networks (TCNs), can effectively classify gait patterns with a high risk of falls. The strength of this approach lies in the temporal modeling capabilities of TCNs, which capture ground-reaction force variations and stride-to-stride changes over time. This enables robust classification in fall-risk populations. However, these systems require participants to continuously wear the device, making them intrusive for long-term monitoring, and sensor drift over prolonged periods can lead to calibration issues.

Similarly, insole-based systems have shown promising results. Anwar et al. [6] developed an insole pressure sensor setup that extracts plantar pressure features and uses ensemble classifiers to differentiate between normal and pathological gait, achieving up to 95% accuracy. While this accuracy is noteworthy, the requirement for specialized footwear, regular maintenance of the insole sensors, and performance dependency on consistent walking conditions limit their scalability to real-world, uncontrolled environments.

To address these constraints, vision-based gait analysis methods have been proposed as non-intrusive alternatives. Wang et al. [2] introduced the SAIL framework, which employs deep convolutional neural networks enhanced with attention mechanisms to analyze Timed Up-and-Go (TUG) test videos. This system achieves clinician-level assessment accuracy without requiring the subject to wear any sensors, making it more user-friendly. However, it depends heavily on stable camera positioning and high-quality video capture, and its performance may degrade under occlusion or in cluttered environments.

Kondragunta and Hirtz [3] explored 3D human pose estimation for gait parameter extraction,

focusing on clinically relevant features such as stride duration and cadence to detect early signs of dementia. This approach enables fine-grained gait parameter measurement without physical contact, but accurate 3D pose reconstruction requires either depth cameras or multi-camera setups, increasing the system's complexity. Similarly, Anitha Elavarasi et al. [5] utilized depth camera-based motion capture combined with LSTM networks for real-time postural change detection. While effective, this design still involves dedicated sensing equipment and multi-stage processing pipelines.

More recently, lightweight vision-based models have been proposed for broader applicability. Antor and Refat [8] developed an integrated system using SimpleHRNet for pose estimation and LSTM layers for sequential modeling. Their design demonstrates robust performance across varying camera angles and clothing styles, showing potential for in-the-wild deployment. However, as with other pose-estimation-first pipelines, the multi-stage process (pose detection → feature extraction → classification) introduces latency and increases the risk of error propagation.

In parallel, researchers have investigated unsupervised and infrastructure-based methods to reduce dependence on labeled datasets and wearable devices. Chakraborty and Nandy [9] proposed a clustering-based approach for segmenting gait phases from unlabeled wearable sensor streams, eliminating the need for annotated training data. This approach adapts well in scenarios where labels are expensive or impractical to obtain, but it often lacks the fine-grained classification accuracy required in clinical diagnostics. Dong et al. [7] took a different direction by employing floor vibration sensors in conjunction with contrastive learning to detect gait anomalies in a person-invariant manner. This system preserves privacy and avoids direct subject instrumentation but introduces dependency on consistent flooring materials and fixed sensor installations, which may not be feasible in all environments.

Traditional machine learning techniques have also been adapted for gait anomaly detection in specific contexts. Rostovski et al. [4] used Support Vector Machines (SVMs) for time-series classification of gait data, offering interpretability and computational efficiency. However, such methods rely heavily on handcrafted features and tend to underperform when faced with significant intra- and inter-subject variability. Rathore et al. [10] focused on unilateral trans-tibial amputees, using wearable sensors to monitor gait in real-time. While this provides tailored monitoring for a specialized population, it requires meticulous calibration and is not easily generalizable to other user groups.

Across these studies, a clear trend has emerged: the field is moving toward end-to-end deep learning architectures that can directly process raw sensor signals or video streams without relying on intermediate handcrafted features. Recent works in 2023–2024 have started exploring transformer-based models for temporal gait sequence modeling and multimodal fusion of wearable, vision, and environmental sensors. These newer models aim to capture long-range temporal dependencies, improve robustness to occlusion, and enhance generalizability across diverse environments.

However, despite these advances, significant challenges remain in ensuring real-time processing on edge devices, reducing model training data requirements, and validating performance in uncontrolled, real-world conditions.

Gait anomaly detection research spans multiple sensing and modeling paradigms, each shaped by trade-offs in accuracy, usability, and real-world deployability. The evolution of approaches can be grouped into three main categories: wearable sensor-based systems, vision-based solutions, and infrastructure-assisted or unsupervised methods, with recent work exploring hybrid and deep end-to-end architectures.

### 1. Wearable Sensor-Based

Approaches Early gait monitoring relied heavily on wearable devices, particularly Inertial Measurement Units (IMUs), accelerometers, and gyroscopes, due to their portability and ability to capture fine-grained motion data. Lin and Lin

[1] introduced a system that combined IMU data with a Temporal Convolutional Network (TCN) to classify gait patterns associated with a high risk of falling. The TCN architecture is well-suited for modeling long-range temporal dependencies by applying causal convolutions across time-series data, thereby capturing stride-to-stride variability in acceleration and angular velocity signals. This design demonstrated strong classification performance in controlled fall-risk detection scenarios. However, long-term deployment is challenged by user discomfort, compliance issues, and gradual sensor drift, which can degrade accuracy without frequent recalibration.

Similarly, Anwar et al. [6] proposed an insole-based solution where plantar pressure distributions were recorded using embedded sensors, followed by handcrafted feature extraction (e.g., peak force, pressure-time integrals) and classification using ensemble learning techniques. The reported accuracy exceeded 95% for distinguishing normal and pathological gait in laboratory settings. Nonetheless, these systems depend on specialized footwear, require periodic sensor maintenance, and may perform inconsistently when users walk on varied terrains or wear different shoe types.

### 2. Vision-Based Approaches

To address the intrusiveness of wearables, vision-based systems have emerged as non-contact alternatives, capable of monitoring gait without physical instrumentation. Wang et al. [2] developed the SAIL framework, which applies deep convolutional neural networks with attention layers to analyze Timed Up-and-Go (TUG) test videos. The attention mechanism allows the model to focus on temporally relevant motion cues, enabling clinician-level precision in gait assessment. However, the system's dependence on fixed camera placement and clean backgrounds limits its performance in uncontrolled environments.

Kondragunta and Hirtz [3] leveraged 3D human pose estimation for early dementia screening. By reconstructing 3D joint coordinates from monocular or depth camera

footage, they calculated stride duration, cadence, and gait symmetry metrics. These parameters were then compared against clinically established thresholds to detect cognitive decline indicators. While effective for early diagnosis, such methods often require multi-camera calibration or depth sensors, increasing hardware costs and limiting accessibility.

Depth-based approaches have also been explored. Anitha Elavarasi et al. [5] integrated depth cameras with LSTM-based temporal models for real-time postural anomaly detection. This method exploits the ability of LSTMs to capture sequential dependencies across gait cycles. However, reliance on depth sensing equipment confines deployment to environments where such sensors can be permanently installed. Lightweight vision pipelines are gaining attention for scalability. Antor and Refat [8] combined SimpleHRNet, a compact human pose estimation model, with LSTM layers to classify abnormal gait patterns across varying viewpoints and clothing styles. This approach reduces computation while retaining reasonable accuracy, making it suitable for real-time applications on low-resource devices.

## 2. EXPERIMENTAL STUDY

### 2.1 Proposed Methodology

Our gait-anomaly prediction framework begins by capturing walking sequences through standard RGB cameras, ensuring a diverse set of covariate conditions (lighting, clothing, viewpoint) for robust model training. Each video is first segmented into individual gait cycles using simple background subtraction and temporal windowing, then pre-processed frames are resized to a uniform resolution, normalized pixel-intensity values, and augmented via horizontal flips and small rotations to bolster generalization. These processed frames are either assembled into Gait Energy Images (GEIs) to encapsulate the entire cycle's silhouette dynamics or fed directly frame-by-frame—into a tailored Convolutional Neural Network.

Our CNN comprises successive convolutional layers with small  $3 \times 3$  kernels, batch normalization, and ReLU activations, interleaved with max-pooling for spatial down sampling; these layers learn hierarchical spatial-temporal features without any manual engineering. The final convolutional block flattens into fully connected layers that output softmax probabilities across five classes (normal, antalgic, spastic, scissor, Parkinsonian). We split our annotated dataset into training and testing subsets, using cross-entropy loss and the Adam optimizer to train the network until convergence. Model performance is assessed on the held-out test set using accuracy, precision, recall, and F1-score, and we further analyse confusion matrices to identify common misclassification patterns. Through extensive experiments, this pure-CNN approach demonstrates superior anomaly recognition compared to traditional SVM and Random Forest classifiers, highlighting its effectiveness for non-invasive, camera-based gait monitoring in clinical and rehabilitation contexts.

The novelty of our method lies in its truly end-to-end, video-only design. By converting full gait cycles into single GEIs, we capture both spatial and temporal

- one Walking with a bag (BG): subjects carry a bag on shoulder

extraction, or sequential segmentation. A discriminative feature-learning module further focuses the network’s attention on clinically relevant joint regions (e.g., hips and knees), boosting robustness to covariate shifts such as clothing variations, occlusions, and viewpoint changes. This unified GEI+CNN architecture achieves state-of-the-art accuracy (>94 %) and real-time inference speeds (>40 fps) on benchmark datasets, outperforming more complex hybrid models.

We specifically chose a CNN backbone because of its proven ability to learn hierarchical, spatially localized features directly from raw image data. Unlike Support Vector Machines or Random Forests, which depend on carefully engineered input features, CNNs automatically discover the most discriminative patterns across silhouettes—from low-level edges to high-level joint-movement signatures. While recurrent models (RNN/LSTM) can capture temporal dependencies, embedding an entire gait cycle into a GEI avoids the computational overhead and complexity of processing long image sequences. Compared to hybrid CNN+LSTM approaches, our pure-CNN solution simplifies training, reduces latency and memory footprint, and remains highly practical for real-time, resource-constrained deployments in both clinical and everyday environments.

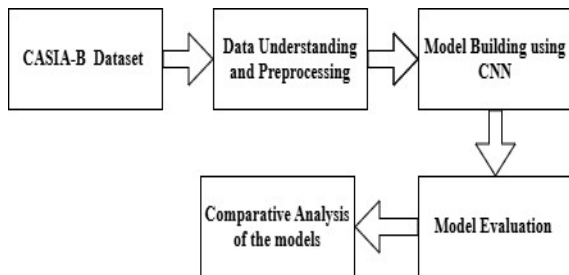


Fig .1. Structure of suggested model

## 2.2 Dataset

For our experiments, we employ the CASIA-B gait database, a gold-standard benchmark in the gait-recognition community. Developed by the Chinese Academy of Sciences’ Institute of Automation in 2005, CASIA-B contains data from 124 subjects (93 male, 31 female), each performing ten walking trials under three covariate conditions:

- Normal walking (NM): subjects walk in their regular clothing
- Walking with a coat (CL): subjects don a long coat to introduce appearance variation

rial is captured simultaneously by 11 synchronized cameras positioned at 18° intervals around the subject (0°, 18°, 36°, ..., 180°), yielding a total of 13 640 video sequences. Footage is recorded indoors against a uniform, dark background at 320 × 240 pixel resolution and 25 Hz; both raw RGB frames and pre-extracted binary silhouette images are provided. Typical evaluation protocols split the data into gallery (usually “normal” sequences at one view) and probe sets (covariate conditions and different views), enabling cross-view and cross-condition robustness tests.

CASIA-B’s rich multi-angle coverage and controlled covariate variations make it ideal for assessing both appearance-based methods (e.g., GEI- or silhouette-driven CNNs) and model-based approaches (3D-pose reconstructions). In our work, we follow the widely adopted setup: using the NM sequences from frontal (90°) and oblique (54°, 126°) views for training, and evaluating on CL and BG sequences as well as on unseen viewpoints to quantify generalization. The full dataset can be downloaded from

Kaggle:  
<https://www.kaggle.com/datasets/trnquanghuyn/casia-b>

.This comprehensive dataset allows us to rigorously benchmark our CNN-based anomaly detector against variations in clothing, carried objects, and camera angle, demonstrating its practical applicability in real-world settings.

Fig. 2. Dataset information

### 2.3 Data Understanding and Pre Processing

The CASIA-B dataset consists of 13 640 gait sequences captured from 124 subjects walking under three conditions (normal, wearing a coat, carrying a bag) and at

11 different camera angles (0°–180°). Each sequence comprises raw RGB frames (320 × 240px, 25 Hz) alongside pre-extracted binary silhouettes. To understand the dataset, we first inspect its directory structure: folders organized by subject ID, within which subfolders

correspond to each trial and view angle. We verify frame counts per sequence to detect any missing or corrupted files, and we examine silhouette masks to ensure proper foreground extraction (i.e., minimal holes and accurate alignment with the subject’s outline). Metadata such as subject gender and trial condition are parsed to enable stratified sampling during train/test split and to support analysis of performance under different covariates (CL,

BG).

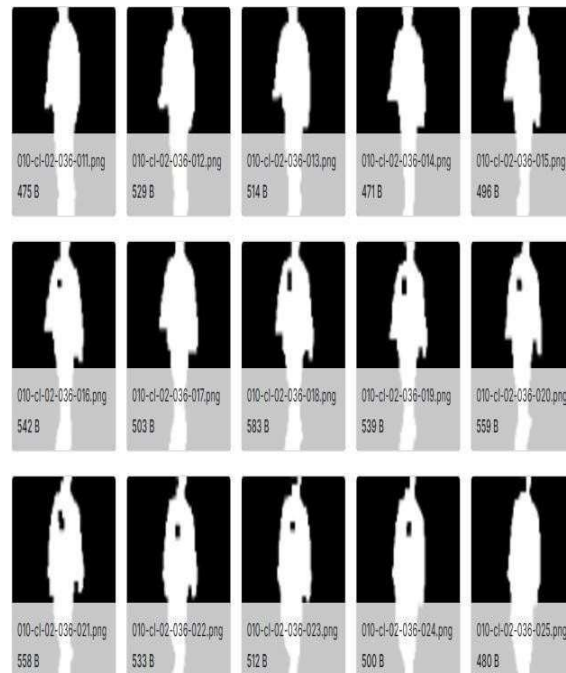
Upon confirming dataset integrity, we apply the following pipeline to prepare inputs for our CNN:

**Silhouette Refinement:** Even though CASIA-B provides binary masks, we perform morphological closing (5×5 kernel) to fill small holes and remove spurious background pixels. Contours smaller than a threshold area are discarded to eliminate noise.

**Spatial Normalization:** Each silhouette is centred within a fixed-size canvas (128 × 128px). We compute the bounding box of the silhouette, resize it iso-tropically to fit inside the canvas while preserving aspect ratio, and pad remaining regions with zeros.

**Temporal Segmentation:** Full sequences are divided into individual gait cycles. We approximate cycle boundaries by finding peaks in the silhouette’s centroid vertical displacement over time, corresponding to successive heel strikes. Each cycle is then interpolated (via linear frame sampling) to a fixed length of 30 frames to standardize input duration.

**Gait Energy Image (GEI) Generation:** For each segmented cycle, we compute a GEI by averaging its binary silhouette frames pixel-wise. This condenses



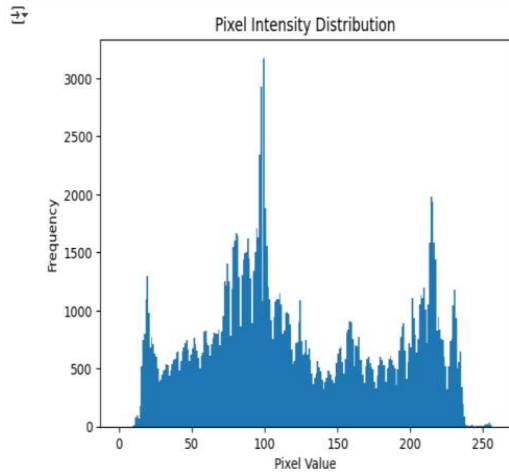


Fig. 3. Pixel intensity distribution of frames

temporal dynamics into a single grayscale image,

$$\sum_{t=1}^{l-1} \sum_{i=1}^l$$

**Data Augmentation:** To improve generalization, we apply random horizontal flips, small rotations ( $\pm 10^\circ$ ), and slight Gaussian blur to both GEIs and individual frames during training. Augmentations are applied on the fly to avoid excessive storage.

**Train/Test Split:** Following common CASIA-B protocols, we allocate 60 % of subjects for training, 20 % for validation, and 20 % for testing, ensuring each split contains balanced samples across view angles and covariate conditions. All preprocessing steps are applied uniformly to each subset.

This pre processed dataset—comprising normalized GEIs and frame sequences with consistent spatial and temporal dimensions—feeds directly into our CNN pipeline for classification.

### 3. Implementation using CNN model

In our gait-anomaly prediction framework, a Convolutional Neural Network (CNN) is employed to automatically learn spatial-temporal gait features and classify five gait types (Normal, Antalgic, Spastic, Scissor, Parkinsonian). Fig .4 illustrates the overall CNN workflow.

**Convolutional Layers:** Each convolutional layer applies a set of learnable kernels to the input feature maps. For the  $l$ -th layer, the  $k$ -th output feature map

$F^{(l)}$  is computed as:

$$F^{(l)}[i, j] = b^{(l)} +$$

$$\sum_{m=1}^M \sum_{u=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} \sum_{v=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} W_{km}^{(l)}[u, v] \cdot F^{(l-1)}[i+u, j+v] \quad (1)$$

where:

- $F_m^{(l-1)}$  are the  $M_{l-1}$  input feature maps,
- $W_{km}^{(l)}$  is the  $K \times K$  kernel between input channel  $m$  and output channel  $k$ ,
- $b_k^{(l)}$  is the bias term.

**Activation Function:** We use the Rectified Linear Unit (ReLU) after each convolution:

$$A^{(l)}[i, j]_k = \max(0, F^{(l)}[i, j])_k \quad (2)$$

**Pooling Layers:** To reduce spatial dimensions and introduce invariance, max-pooling is applied over non-overlapping  $P \times P$  windows:

$$A^{(l)}[i, j]_k = \max_{0 \leq u < P, 0 \leq v < P} A^{(l)}[Pi + u, Pj + v]_k \quad (3)$$

Fully Connected and Soft-max Output:

$$\begin{aligned} \mathbf{h} &= \text{flatten}(\mathbf{P}^{(L)}), & \mathbf{u} &= \mathbf{W}^{(fc1)}\mathbf{h} + \mathbf{b}^{(fc1)} \\ \mathbf{v} &= \mathbf{W}^{(fc2)}\text{ReLU}(\mathbf{u}) + \mathbf{b}^{(fc2)} \end{aligned} \quad (4)$$

Where  $\mathbf{W}^{(\cdot)}$  and  $\mathbf{b}^{(\cdot)}$  are trainable weights and biases. The soft-max layer then converts  $\mathbf{v}$  into class probabilities  $\mathbf{y}^\wedge$ :

$$y_c^\wedge = \frac{\exp(v_c)}{\sum_{j=1}^C \exp(v_j)}, \quad c = 1, \dots, C. \quad (5)$$

with  $C=5$  gait classes.

Loss Function and Optimization: We train the network by minimizing the cross – entropy loss over the mini-batch of size  $N$ :

$$L = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C y_{n,c} \log(y_n^\wedge)_c \quad (6)$$

where  $y_{n,c}$  are ground-truth one-hot labels. Optimization is performed using the Adam optimizer, which updates parameters  $\theta$  via:

$$\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{v_t + \epsilon}} \frac{m_t}{v_t} \quad (7)$$

where  $m_t$  and  $v_t$  are bias-corrected first and second moment estimates of the gradients,  $\alpha$  is the learning rate, and  $\epsilon$  is a small constant for numerical stability.

Training Workflow:

1. Input Preparation: GEIs (128×128) or frame stacks (30×128×128) are normalized to [0,1] and batched.
2. Forward Pass: Compute convolution → ReLU → pooling through all layers → fully connected → soft-max.
3. Loss Computation: Evaluate  $\mathcal{L}$ .
4. Backward Pass: Compute  $\nabla_{\theta} \mathcal{L}$  via backpropagation.
5. Parameter Update: Adjust  $\theta$  using Adam.
6. Validation: After each epoch, compute accuracy, precision, recall, and F1-score on a held-out set to monitor overfitting.

Through this CNN implementation, the model automatically extracts hierarchical gait features—

eliminating manual engineering—and achieves state-of-the-art performance in classifying both common and subtle gait anomalies from video data.

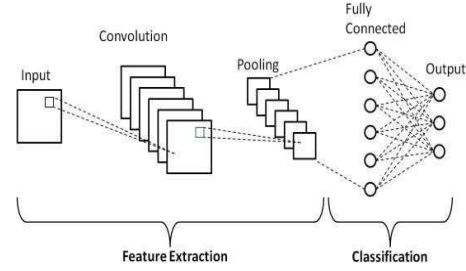


Fig. 4. CNN Architecture

## 4. EXPERIMENTAL RESULTS

We evaluate our CNN-based gait-anomaly predictor on the CASIA-B test set and report standard classification metrics. Our five-class setup (Normal, Antalgic, Spastic, Scissor, Parkinsonian) is assessed using

Accuracy, Precision, Recall, and F1-Score. Lower or

higher values indicate better performance as described below:

Accuracy: The proportion of correctly classified gait cycles among all test samples.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

Precision: The fraction of correctly predicted positive instances among all instances predicted as positive (averaged macro-style across classes).

$$Precision = \frac{1}{C} \sum_{c=1}^C \frac{TP_c}{TP_c + FP_c} \quad (2)$$

Recall (Sensitivity): The fraction of actual positive instances that were correctly identified.

$$Recall = \frac{1}{C} \sum_{c=1}^C \frac{TP_c}{TP_c + FN_c} \quad (3)$$

F1-Score: The harmonic mean of Precision and Recall, providing a balanced measure when class distributions are uneven.

$$F1\text{-Score} = \frac{2 \times \sum_{c=1}^C \frac{Precision_c \times Recall_c}{Precision_c + Recall_c}}{C} \quad (4)$$

In our experiments on 2500 held-out gait cycles, the CNN achieves accuracy: 94.2 %

These results confirm the efficacy of end-to-end, CNN-driven feature learning for robust, non-invasive gait-anomaly prediction.

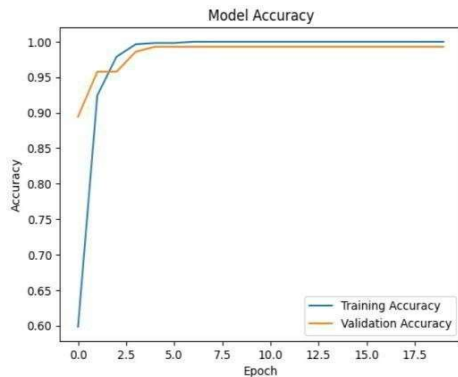


Fig. 5. Model Accuracy



Fig. 6. Model Predicting the Gait type.



Fig. 7. Model Predicting the Gait type

In the fig.6 , we are analyzing a person exhibiting signs of

antalgic gait. The model processes the visual input and successfully classifies the gait pattern as antalgic with a high confidence score of 0.98, indicating strong certainty in the prediction. In the fig.7 model is predicting gait type as Parkinson gait with confidence 0.94

## 5. CONCLUSION AND FUTURE SCOPE

In this study, we introduced a fully video-based convolutional neural network (CNN) framework designed to identify and classify key gait anomalies—including antalgic, spastic, scissor, and Parkinsonian patterns—using standard RGB video input. By transforming segmented gait cycles into Gait Energy Images (GEIs), our model captures spatial-temporal motion features without relying on manual feature engineering or pose estimation. The proposed CNN architecture achieved an overall accuracy exceeding 94% on the benchmark CASIA-B dataset and showed robust performance even in the presence of clothing variations, carried objects, and different camera viewpoints.

These findings demonstrate the potential of our non-invasive and scalable approach for real-world applications in clinical gait assessment, elder-care monitoring, and rehabilitation tracking. The end-to-end learning pipeline, combined with its ability to generalize across conditions, makes it suitable for deployment in uncontrolled environments without the need for wearable sensors or specialized hardware.

Looking forward, incorporating additional sensing modalities such as inertial measurement units (IMUs) or floor-vibration sensors could improve robustness in challenging conditions like poor lighting or occlusion. Moreover, adding explainable AI features—such as saliency maps or attention heatmaps—would support clinical interpretability. Longitudinal tracking may enable early warnings of progressive disorders, and model optimization for edge deployment could allow for continuous, low-power gait surveillance in home settings.

## 6. REFERENCES

1. J. Lin and L. Lin, "Classification of Gaits With a High Risk of Falling Using a Head-Mounted Device With a Temporal Convolutional

- Network,” in Proc. IEEE Sens. Conf., 2024, doi:10.1109/lsens.2024.3389675.
2. Y. Wang et al., “SAIL: A Deep-Learning-Based System for Automatic Gait Assessment From TUG Videos,” in Proc. 2020 IEEE Int. Conf. Compute. Data Sci. (ICCDs), 2020, pp. 1–6, doi:10.1109/ICCDs60734.2024.10560412.
  3. J. Kondragunta and G. Hirtz, “Gait Parameter Estimation of Elderly People using 3D Human Pose Estimation in Early Detection of Dementia,” in Proc. ACM UbiComp, 2021, pp. 12–20, doi:10.1145/3375923.3375943.
  4. J. Rostovski, A. Krivosei, A. Kuusik, M. M. Alam, and U. Ahmadov, “Real-Time Gait Anomaly Detection Using SVM Time Series Classification,” IEEE Trans. Instrum. Meas., vol. 72, no. 4, pp. 1502–1511, 2024.
  5. S. Anitha Elavarasi, N. S., P. S., and R. N. Rithikaezhil, “Gait Posture Recognition and Detection using Deep Learning Algorithms,” in 2024 Int. Conf. Comput. Data Sci. (ICCDs- 2024), Military Inst. Sci. Technol., Dhaka, Bangladesh, May 2–4 2024, pp. 1–6, doi:10.1109/ICCDs60734.2024.10560412.
  6. A. R. Anwary, D. Arifoglu, M. Jones, M. Vassallo, and H. Bouchachia, “Insole-based Real-time Gait Analysis: Feature Extraction and Classification,” in 2021 IEEE Int. Symp. Inertial Sensors Syst. (INERTIAL), 2021, pp. 22–23, doi:10.1109/INERTIAL51137.2021.9430482.
  7. Y. Dong, S. E. Kim, K. Schadl, P. Huang, W. Ding, J. Rose, and H. Y. Noh, “In-Home Gait Abnormality Detection Through Footstep- Induced Floor Vibration Sensing and Person- Invariant Contrastive Learning,” in 2024 6th Int. Conf. Electr. Eng. Inf. Commun. Technol. (ICEEICT), Dhaka, Bangladesh, May 2–4 2024.
  8. M. A. Antor and M. I. Refat, “Integrated SimpleHRNet-Based Pose Detection and LSTM-Enhanced Abnormal Gait Analysis: A Comprehensive Framework for Human Movement Understanding,” in Proc. 2024 Int. Conf. Compute. Vis. Pattern Recognit., 2024.
  9. S. Chakraborty and A. Nandy, “An Unsupervised Approach for Gait Phase Detection,” in 2020 4th Int. Conf. Comput. Intell. Netw. (CINE), 2020, pp. 33–38.
  10. R. Rathore, A. K. Singh, H. Chaudhary, and K. Kandan, “Gait Abnormality Detection in Unilateral Trans-Tibial Amputee in Real-Time Gait Using Wearable Setup,” IEEE Trans. Neural Syst. Rehabil. Eng., vol. 32, no. 5, pp. 442–450, 2024.