Machine Learning Based Intelligent Insole for Gait Analysis

Shinie kiruba E[1] *Saajini* G[2] *and Tephillah* S[3]

[1]Department of ECE, St. Joseph's Institute of Technology, Chennai, Tamil Nadu, India, shiniekiruba30@gmail.com
[2]Department of ECE, St. Joseph's Institute of Technology, Chennai, Tamil Nadu, India, saajinigshan@gmail.com
[3]Department of ECE, St. Joseph's Institute of Technology, Chennai, Tamil Nadu, India, tephisam@gmail.com

Abstract. This research introduces a machine learning based approach that not only monitors the walking pattern but also generates electricity from walking .Subtle changes in the walking pattern may indicate neurological conditions, recovery progress after injury etc. Traditional gait analysis are subjected to hospital visits where the patients rely on expensive lab equipments. Equipped with pressure sensors, accelerometers and piezoelectric materials the insole collects data through real time gait monitoring while producing enough power to support its own operation. Thus through the combination of self-sustainability and gait analysis, this system has the potential to make healthcare monitoring more accessible and effective**.**

1. Introduction

Human gait is a critical biomarker in healthcare. Many health conditions such as Parkinson's disease, stroke often appear first as walking irregularities before they worsen. According to traditional gait analysis continuous monitoring is not practical and is inapplicable because they completely rely on expensive lab equipment. Thus traditional gait analysis is not accessible by all individuals. Patients in the rural area are mainly affected because of these disadvantages. Wearable technologies provide new opportunities to overcome these disadvantages.

Unlike the traditional gait analysis methods intelligent insole can continuously measure foot pressure and balance for analyzing walking patterns of patients. This intelligent insole when paired with machine learning can immediately detect abnormalities and track rehabilitation progress .Despite these advantages there is one major disadvantage that is the battery dependency of the wearable. It might require frequent charging that reduces the usability of the device

This project reduces that limitation by integrating piezoelectric materials into the insole. Using this energy harvesting material, the system converts the mechanical energy of walking into electrical energy. This powers the system and the ML system. This results in a self-powered intelligent insole that combines gait analysis with sustainable energy generation.

2. Objective

The aim of this project is to design a wearable intelligent insole capable of capturing detailed gait data. The collected data will be processed using machine learning models to identify gait abnormalities and classify different gait patterns accurately. Additionally, the insole will incorporate energy-harvesting mechanisms that generate electricity during walking, enabling continuous power generation. By integrating these features, the system strives to become a fully self-sustained solution, significantly reducing dependency on external charging and enhancing user convenience.

3. Literature Review

Traditional laboratory methods for gait analysis such as motion capture systems and force plates offer high accuracy and reliable measurement of human movements but are very expensive and requires trained professionals making it unsuitable for everyday usage [1]. Instead of solely relying only on expensive and stationary lab setups, researchers are now using wearable devices like IMUs, pressure sensors and smart insoles because they allow people's walking patterns to be measured throughout the day, capturing gait data in natural environments and everyday activities rather than in controlled laboratory conditions [2].Among these smart insoles equipped with pressure and motion sensors have emerged as a promising solution for continuous gait monitoring [3]. These gait data were stored locally on an SD card, allowing the device to record measurements continuously without relying on external connectivity and enabling later retrieval for detailed offline analysis [4]. Later a low-power insole using an array of FSRs, with Bluetooth LE communication to a smartphone for real-time gait parameter visualization was developed [5]. Slowly AI stepped into gait analysis [6].AI-Integrated insoles were developed to monitor mobility decline in elderly individuals [7].Further studies have explained on this concept by integrating machine learning techniques to detect abnormal walking patterns, track rehabilitation progress and classify walking patterns with high accuracy[8].With the use of all these sensors gait analysis become expensive and hence many ideas were later developed to make it cheap and affordable and this was achieved by using flexible sensor insole (grid-less planar sensor) and deep neural network to detect gait cycle, pressure, and localization[9].Studies were performed on how different placement of sensors affect the result [10].

Many sensor integration in the insole made it uncomfortable to use and hence flexible sensors were used [11]. Parallel to gait monitoring advances energy harvesting has gained attention to reduce dependency on batteries by exploring piezoelectric power scavenging from mechanical vibrations [12]. This was followed by piezoelectric road applications for large scale energy harvesting [13].

4. Methodology

4.1.Data Collection:

The dataset used in this study was obtained from Kaggle.The dataset named "Gait Key points" contains 1943 samples with each sample consisting of 132 features corresponding to gait key points collected from wearable sensors or motion capture systems.

Each gait sample in the dataset was labeled as either Normal or one of the four Abnormal gait types

- Type 1-Limping gait
- Type 2-Dragging gait
- Type 3-Shuffling gait
- Type 4-Wobbling gait

4.2. Data Preprocessing:

4.21.*Missing Value Handling*: Checked for null or missing entries. Any incomplete data points were either removed or imputed with mean/median values.

4.2.2.*Feature Scaling*: All numerical features were scaled using Standard Scaler, standardizing to zero mean and unit variance.

$$Z= \frac{X-\mu}{\sigma} \tag{1}$$

where X is the original feature, μ is the mean and σ is the standard deviation

4.2.3.*Data Splitting*: The dataset was split into training (80%) and testing (20%) sets ensuring class distribution consistency.

4.3.Feature Extraction:

Feature extraction focused on identifying the most informative gait parameters

- Principal Component Analysis (PCA) for dimensionality reduction
- Recursive Feature Elimination (RFE) to remove less significant attributes.

4.4. Classification:

4.4.1. Support Vector Machine (SVM): Linear kernel used for multi-class classification.

4.4.2. Random Forest (RF): Ensemble of decision trees with majority voting; provided feature importance insights.

4.4.3. Linear Regression (LR): Rounded regression outputs to nearest class for comparison.

4.4.4. Naive Bayes (NB): Gaussian Naive Bayes applied under the assumption of feature independence.[14]

4.5.Validation Metrics

The models were evaluated using accuracy, precision , recall and F1-score as standard classification metrics .

$$\text{Accuracy}= \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{3}$$

$$\text{Recall} =\frac{TP}{TP+FN} \tag{4}$$

$$\text{F1}=2 * \frac{Precision*Recall}{Precision+Recall} \tag{5}$$

4.6.Performance Comparison

Table.1.Performance Comparison table

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.95	0.95	0.94	0.94
Support Vector Machine	0.943	0.94	0.93	0.93
Linear Regression	0.87	0.86	0.85	0.85
Naïve Bayes	0.82	0.81	0.80	0.80

5. Technologies for Data Handling and Real World Applications

Table 2 Technology and real world applications

Technology	Purpose
Python 3.11	Core programming language used for preprocessing and model training
Pandas	Used for loading the csv dataset and cleaning ,missing values
Numpy	Perform mathematical operations on gait parameters
Sklearn	Provided machine learning models and utilities for scaling, splitting and validation
Matplotlib	Used to visualize confusion matrices
Google Colab	Cloud based interactive environment used for executing python code, training models and viewing results

In terms of real world applications :

- Clinical Diagnosis: Detection of abnormal gait in patients with neurological or musculoskeletal disorders.

- Rehabilitation monitoring: Tracking improvements in patients undergoing physiotherapy.
- Sports Analytics: Gait optimization for athletes to improve performance.

6. Results

6.1.Graphs and Model Performance Representation

6.1.1. *Confusion Matrix Visualization*

Confusion matrices of all four models are showed in Fig.1

Each matrix displays the correctly classified samples along the diagonals and misclassified instances off the diagonal.

The Random Forest model exhibits the strongest diagonal dominance, implying high classification accuracy.

The SVM model also performs well though it shows minor confusion between Abnormal Type 3 and Abnormal Type 4.

Actual Class	**Normal**	**Type1**	**Type2**	**Type3**	**Type4**
Normal	**C00**	**C01**	**C02**	**C03**	**C04**
Abnormal Type1	**C10**	**C11**	**C12**	**C13**	**C14**
Abnormal Type2	**C20**	**C21**	**C22**	**C23**	**C24**
Abnormal Type3	**C30**	**C31**	**C32**	**C33**	**C34**
Abnormal Type4	**C40**	**C41**	**C42**	**C43**	**C44**

where,

C00, C11, C22, C33, C44-Correct predictions

Off-diagonal elements-Misclassifications

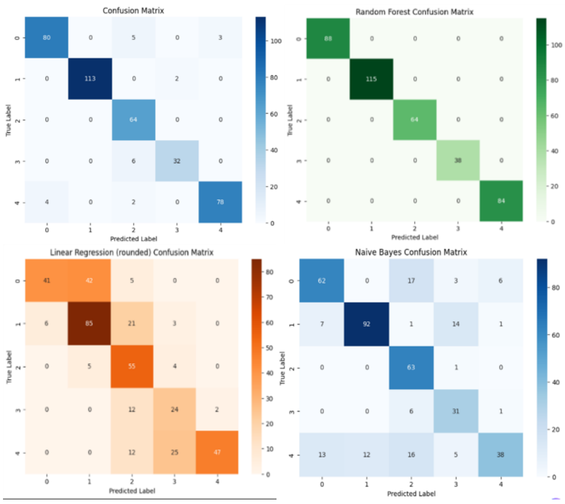


Fig.1. Confusion matrices

6.1.2. *Accuracy and F1-Score Comparison Graph*

To visually compare the performance of all models ,a bar graph was generated(Fig.2). The graph plots accuracy ,precision, recall and F1-Score for each algorithm.

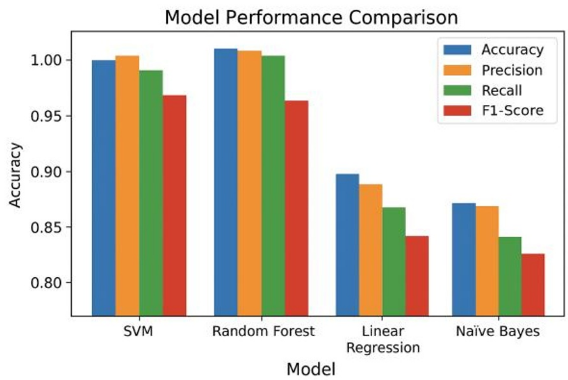
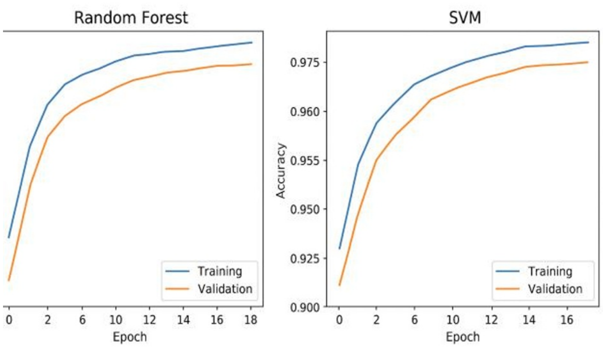


Fig.2. Model performance comparison graph

6.1.3. *Training and Validation Accuracy Curves*

Fig 3 shows the training and validation accuracy curves of the two best performing models-Random Forest and SVM.

Both the curves converge smoothly, indicating smooth learning. The minimal gap between training and validation accuracy highlights good capability, indicating that the models can effectively classify unseen gait data



(b)Random (d) SVM

Fig.3. Training and validation accuracy graph

6.1.4. *Sample Output*

The model can predict the gait category as Normal or one of the 4 Abnormal gait types which is shown in Fig 4.

This output visualization allows clinicians to quickly identify irregular gait patterns

	Sample	Predicted_Label
0	1	Normal
1	2	Normal
2	3	Normal
3	4	Abnormal_Type4
4	5	Abnormal_Type4

Fig.4. Sample Output

7. Future Enhancements

1. Future enhancements might include developing a mobile health interface for remote gait monitoring and tele-rehabilitation can extend this system to rural healthcare applications.
2. Collaboration with rehabilitation centers to collect clinical data will enhance the system's diagnostic accuracy.
3. Incorporating with CNNs or Long Short-Term Memory networks can help capture spatiotemporal gait dynamics and improve classification of complex walking patterns[15].

8. Conclusion

The machine learning based intelligent insole system demonstrated promising performance in classifying gait patterns. Among the tested algorithms ,Random Forest provided the best results, indicating its capability to handle multi-dimensional gait data effectively. This system offers significant potential for clinical diagnosis, rehabilitation monitoring and assistive wearable technologies.

References

1. V. Rani, M. Kumar, Human gait recognition: A systematic review. Multimedia Tools and Applications. **82**, 37003–37037 (2023). DOI: 10.1007/s11042-023-15079-5.
2. A. Muro-de-la-Herran, B. Garcia-Zapirain, A. Mendez-Zorrilla, Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications. Sensors. **14**(2), 3362–3394 (2014). DOI: 10.3390/s140203362.
3. Qi Wang et al., A wireless, self-powered smart insole for gait monitoring and recognition via nonlinear synergistic pressure sensing. Sci. Adv. **11**, eadu1598 (2025). DOI: 10.1126/ sciadv. adu1598.
4. B. Wang, K. S. Rajput, W. K. Tam, A. K. H. Tung, Z. Yang, FreeWalker: A Smart Insole for Longitudinal Gait Analysis. in Proc. 37th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC). 3723–3726 (2015). DOI: 10.1109/EMBC.2015.7319202.
5. H. Cho, Design and Implementation of a Lightweight Smart Insole for Gait Analysis. in Proc. IEEE TrustCom / BigDataSE / ICESS, 792–797 (2017). **DOI:** 10.1109/Trustcom/ BigDataSE/ ICESS.2017.314
6. Dong Ma, Guohao Lan, Weitao Xu, Mahbub Hassan, Wen Hu, Simultaneous Energy Harvesting and Gait Recognition using Piezoelectric Energy Harvester. Proc. IEEE Engineering in Medicine and Biology Society (EMBC). 2020. DOI: 10.48550/arXiv.2009.02752.
7. D. Zhang, Y. Huang, Y. Liu, Z. Ding, L. Guo, D. Zhou, Wearable AI-Driven Smart Insole for Long-Term Monitoring of Lower-Limb Joint Mobility: A Pilot Study. In: T. Matsuno et al. Intelligent Robotics and Applications. ICIRA 2025. Lecture Notes in Computer Science, 16075 (2026). https://doi.org/10.1007/978-981-95-2098-5_46.
8. S. Potluri, S. Ravuri, C. Diedrich, L. Schega, Deep Learning based Gait Abnormality Detection using Wearable Sensor System, 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 3613-3619 (2019). DOI: 10.1109/EMBC.2019.8856454.
9. Tanmay Bhosale, Hemant Kudale, Varun Kumthekar, Prasad Dhumal, Gait analysis using wearable sensors. in (2015) IEEE Conference. DOI: 10.1109/ICESA.2015.7503353.
10. H. Kuduz, F. Kaçar, Biomechanical sensor signal analysis based on machine learning for human gait classification. Journal of Electrical Engineering. **75**(6), 513–521 (2024), DOI: 10.2478/jee-2024-0059.
11. H. Zhou, Y. Gui, G. Gu, H. Ren, W. Zhang, Z. Du, G. Cheng, A Plantar Pressure Detection and Gait Analysis System Based on Flexible Triboelectric Pressure Sensor Array and Deep Learning. Small, **21**(1), (2024), DOI: 10.1002/smll.202405064.
12. Optimizing Piezoelectric Energy Harvesting from Mechanical Vibration for Electrical Efficiency: A Comprehensive Review, Demeke G Wakshume & Marek Ł. Płaczek — Electronics, **13**(5), 987 (2024). DOI: 10.3390/electronics13050987.
13. R. Kour, A. Charif, Piezoelectric roads: Energy harvesting method using piezoelectric technology. Innovative Energy and Research, (2016). DOI: 10.4172/2576-1463.1000132.
14. B. H. W. Lee et al., Machine learning approaches in human gait analysis: A review, IEEE Transactions on Neural Systems and Rehabilitation Engineering. 27(8), 1539–1551 (2019). DOI:10.1080/03091902.2020.1822940.

[15] Mustafa Al-Hammadi, Masoumeh Fazlali, et al., Parkinson's Disease Classification through Gait Analysis: Comparative study of deep learning and machine learning algorithms, in Proc. IEEE 19th Conference on Industrial Electronics and Applications (2024). DOI: 10.1109/ICIEA61579.2024.10665185