2026 IEEE International Symposium on Circuits and Systems (ISCAS) | 979-8-3315-7769-8/26/$31.00 ©2026 IEEE | DOI: 10.1109/ISCAS66217.2026.11562098

Multimodal Smart Insole with Crossbar Crosstalk Compensation for Fall-Risk Prediction

Junjian Chi, Zihuan Zhang, Qingyu Zhang, Andreas Demosthenous and Yu Wu
Department of Electronic and Electrical Engineering, University College London, Torrington Place, London WC1E 7JE, UK
Email: junjian.chi.22@ucl.ac.uk; zihuan.zhang.24@ucl.ac.uk; yu.wu.09@ucl.ac.uk

***Abstract* — Gait analysis has always relied on latent indicators that can provide early warnings about underlying health conditions. Fall-risk prediction is one of the most important applications. In this paper, we present a portable smart insole system that couples a 253-sensor plantar pressure array with a 6-DoF IMU and an end-to-end deep learning fall-risk prediction pipeline. A double-frame sensing scheme expands the sensor's dynamic range under different pressure loadings. To suppress sneak-path crosstalk, a compact U-Net (33×15 input) reconstructs a crosstalk-free matrix from a cross-bar array, achieving R^2 = 0.9307. The clean pressure map and IMU signals are input into a 3D-CNN + Transformer encoder for joint pose regression (13×3) and fall-risk prediction, achieving accuracy of up to 91%. With this indicator, we can determine whether the cumulative probability of a person falling is steadily increasing over time. This further suggests that their physical health status may warrant additional examination, allowing the system to issue an early warning to the user. Fall itself is instantaneous, but the factors that lead to a fall are issues worth monitoring over the long term.**

Keywords—Gait analysis, Wearable insole system, Fall-risk prediction, Multimodal deep learning, Crossbar-array crosstalk compensation

I. Introduction

Falling is one of the primary causes of global injury events, especially in the elderly, leading to many physical damages and dramatic psychological consequences, which reduce the independence of the person [1]. It is typically an instantaneous event, leaving little chance to react and intervene. However, most fall protection systems are post-hoc alarm systems that trigger only after a fall has occurred [2], [3]. The development of a pre-fall alert system should be brought to attention. Gait, as a behavioural and postural characteristic of human locomotion, reflects subtle pre-fall perturbations and latent health changes. Previous research has shown that gait biomarkers, such as gait speed and variability, are discriminative indicators of increased fall risk [4]. Currently, various portable gait analysis systems include wireless non-wearable systems, phone-based non-wearable systems, and fully integrated wearable systems [5]. As non-wearable systems are not suitable for long-term use, integrated wearable systems are a better option for long-term fall risk assessment.

Wearable technologies that integrate pressure and inertial sensors in the form of insoles are typically used for analysing gait characteristics. To capture high-resolution plantar pressure inputs, a cross-bar pressure sensor array is commonly adapted on the insole [6-8] due to its simple structure, cost-efficient fabrication, and programmable sensing geometry [9]. However, the sneak path, which leads to significant crosstalk, is a major issue in this type of scheme. Since rows and columns interconnect adjacent units, activating a single sensor cell can influence other cells in the same row and column through parallel pathways in the array, leading to measurement artefacts, especially in multi-touch scenarios [10]. There are several hardware solutions, including diode isolation, MOSFET-based active matrix, and the zero-potential method (ZPM) [11], [12], [13]. Diode isolation connects the diode with the resistive sensor to enforce unidirectional conduction. A MOSFET-based active matrix uses MOSFETs as switches to put unselected sensors into a high-impedance state. The ZPM method actively drives non-selected lines to equal potential to suppress parallel leakage. While these methods can provide an undistorted sensor matrix, the additional components and routing can increase costs and fabrication complexity and raise the overall system's power consumption.

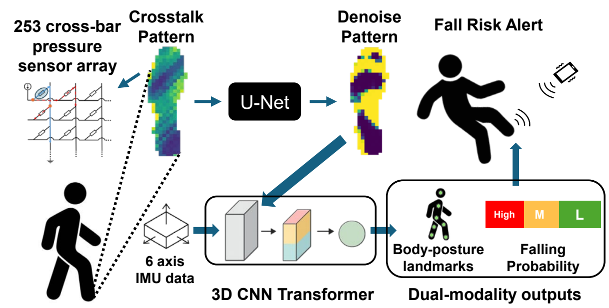

Fig. 1. Overview of portable smart insole system for fall prediction.

In this paper, a portable insole system integrating 253 pressure sensors and inertial measurement unit sensors (IMU) is described. The hardware is improved over our previous version described in [14], featuring a compact yet powerful analogue front-end that offers high SNR, and dynamic range. As shown in Fig. 1, we treat the crosstalk issue as an image task and apply an end-to-end restoration network (a U-Net) to learn to overcome the sneak-path effects. At the application level, "fall prediction", we designed a multimodal 3D CNN with a Transformer encoder that takes both the pressure and IMU data combined to predict both lower-body joint movements and the probability of falling. The model achieves 91% accuracy in fall prediction.

The rest of this paper is organised as follows: Section II discusses the hardware architecture and the U-Net design and implementation for addressing crosstalk. Section III describes the details of the multimodal 3D CNN with a Transformer encoder and evaluates the model results. Conclusions are in Section IV.

This work was supported by the Rosetrees Trust [UCL-IHE-HA\1], the Engineering and Physical Sciences Research Council (EPSRC) [grant number UKRI1493], and the UCL NNTI Joint Lab fund.

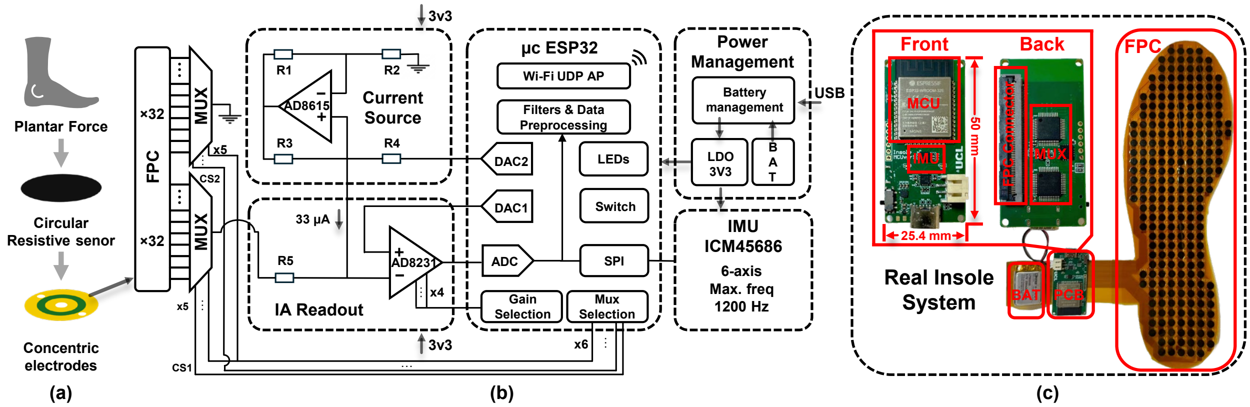

Fig. 2. (a) Sensor stacks. (b) Hardware architecture. (c) Final insole system.

II. System Architecture and Design

The insole's cross-bar pressure sensor array is built onto a flexible printed circuit board (FPC), with the cross-bar pattern etched onto the FPC. The crossing node consists of two concentric circular electrodes, as shown in Fig. 2 (a), connecting the rows and columns to a piezoresistive pressure sensor (Velostat). When the sensors are pressed against the circular electrodes, their resistance change is read by the hardware system, as shown in Fig. 2 (b). The challenges for measurement lie in 1) the piezoresistive sensors' resistance has a non-linear characteristic across the targeted plantar pressure range with a wide dynamic range, and 2) the presence of the mentioned sneak path issue inherent in the cross-bar sensor array.

A. Double-Frame Sensing

Here, we highlight a double-frame sensing scheme in our system to overcome the challenges. As shown in Fig.2 (b), the sensor is excited by a Howland current source with voltage readout using an instrumentation amplifier (IA). Having the four resistors of the Howland current source equal to R_1 = 100 kΩ, the output current is:

$$I_{\text{OUT}} = \frac{V_{DAC_2}}{R_1} \tag{1}$$

where V_{DAC_2} is the output of DAC_2, and I_{OUT} at its maximum is $33uA$. An R5 = 15 kΩ is inserted in series with the sensor to establish a minimum common-mode voltage range for the IA. By sweeping DAC_2 while keep DAC_1 at a fixed value, it is possible to calibrate sensor variations using a Look-Up-Table method. The final IA output is then:

$$V_{\text{out}} = \text{Gain} \cdot \left(V_{DAC_1} - \frac{V_{DAC_2}}{R_1}(15k + R_{\text{sensor}}) \right) \tag{2}$$

The sensor under low and high pressure can vary in the ranges of tens of kΩ, decreasing non-linearly with increasing pressure. We adapted two frames with complementary resistance ranges: 1) Frame 1: V_{DAC_1} = 1.6V, Gain = 1X, covering the whole pressure range with lower sensitivity. 2) Frame 2: V_{DAC_1} = 0.8V with higher gain starting from 8X, increasing sensitivity in the high-pressure (low-resistance) region, and clipping the low-pressure (high-resistance) region. The first frame provides a wide measurable range, and the second one provides higher resolution. We overlay the two frames via scaled addition to maximise effective resolution across the entire insole region, often under different pressure loadings.

In addition, we integrate a 6 Degree-of-Freedom (DOF) IMU ICM-45686 to capture 3-axis acceleration and 3-axis angular velocity communicated by ESP32-WROOM-32E, providing complementary information on top of the pressure data. The digital controls and wireless transmission scheme follow our previous design [14], with each payload including two frames of 253 pressure matrices and 6-DOF IMU data; the transmission rate decreases to 30 Hz from 60 Hz.

B. U-Net for Cross-talk

Another major issue still exists in this setup: the sneak path in the cross-bar array. Reconstructing a crosstalk-free pressure map from a cross-bar resistance matrix mathematically is solving an ill-posed inverse problem: multiple sneak paths make the inverse mapping solution non-unique. The U-Net has proven effective for inverse tasks such as sparse-view CT [15], even with limited datasets. Therefore, we recast the task as an image-to-image reconstruction problem and adopt a U-Net to learn the crosstalk-to-crosstalk-free mapping end-to-end.

As illustrated in Fig. 3 (a), our U-Net model follows a design of 4 down-sampling stages and is tailored to a small input of 33×15 as the size of our pressure heatmap. The encoder on the left side applies 3×3 convolutions before each down-sampling stage to extract increasingly abstract features while maintaining the spatial size. At the down-sampling stage, a 2×2 max pooling layer will be applied to reduce computation while enlarging the effective receptive field. The input 33×15 heatmap will undergo three such pooling layers, it will be gradually reduced to a size of 16×7, 8×3, and finally 4×1. As the fourth 2×2 pooling would collapse the width to zero, we replace it with a 1×1 operation. From the bottom, the decoder mirrors the encoder with a sequence of 2×2 up-samplings to restore the original resolution of 33×15 progressively. After each up-sampling, we concatenate the corresponding encoder feature map along the channel dimension before applying 3×3 convolutions again.

The skip connections as shown are significant because they reinject high-resolution and low-level details, such as heatmap boundaries, that were lost during the pooling process. A final 1×1 convolution acts as a pixel-wise classifier,

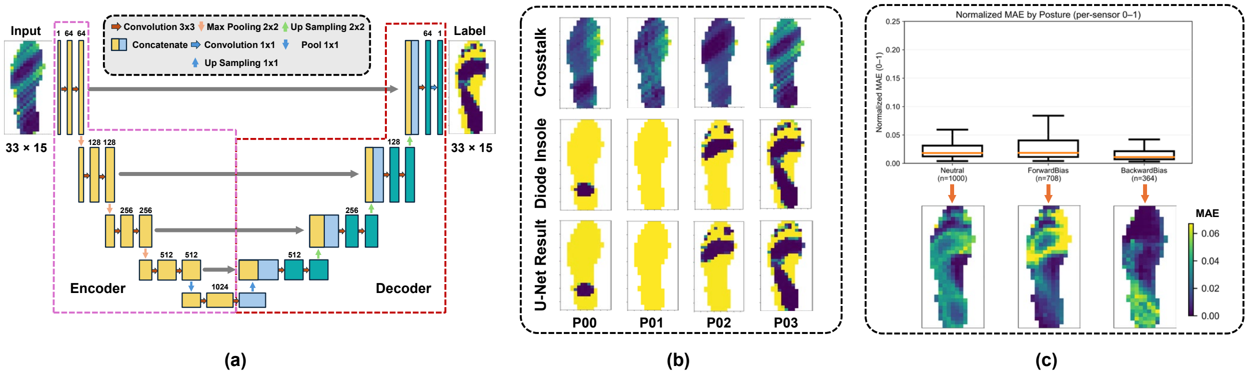

Fig. 3. (a) Crosstalk U-Net model. (b) Visualization examples of crosstalk input, diode insole label and U-net result for four postures. (c) Normalised MAE for P01, P02 and P03 (top) and per-pixel normalised-MAE heatmaps (bottom).

mapping the fused features at 33×15 back to the desired number of output channels without altering spatial size.

The procedure of data collection is done by stacking the standard insole (with crosstalk) with a crosstalk-free diode insole (diode-connected between each sensor and its column node) so that both the crosstalk and crosstalk-free pressure heatmaps can be obtained simultaneously. Under the same pressure load, the diode insole suppresses sneak paths and provides a clean ground truth, whereas the standard insole exhibits crosstalk and serves as the input to the U-Net. Specifically, we define four foot postures: P00: Backwards Bias, P01: No-load, P02: Forward Bias and P03: Neutral (whole-foot stand). The two data streams collected are temporally aligned, approximately 12k pairs of datasets, which are randomly split into 2k pairs for testing, and the remaining 10k pairs are randomly divided into training and validation sets with an 80/20 ratio.

TABLE I CROSSTALK U-NET MODEL RESULT

	MSE	MAE	RMSE	R^2	PSNR (dB)
P01	**0.012966**	**0.030707**	**0.113867**	**0.9430**	**18.87**
P02	**0.020948**	**0.039433**	**0.144734**	**0.8867**	**16.79**
P03	**0.017744**	**0.034271**	**0.133205**	**0.8994**	**17.51**
Total	**0.014184**	**0.029854**	**0.119097**	**0.9307**	**18.48**

As shown in Fig. 3 (b), we have selected and visualised the labelled and predicted pressure heatmaps for four postures. The model essentially recovers the shape and the pressure distribution of the plantar contact and can recognise the places where no force is applied. Fig. 3(c) illustrates a boxplot of the normalised MAE of posture P00, P02 and P03 and the corresponding MAE per-pixel heatmap. The error distribution for P00 is the most concentrated, with a low upper quartile; P02 has the highest median with a long tail, indicating that it is more difficult for our model to reconstruct the forefoot pressure. Evaluation metrics for each posture and the overall are presented in Table 1. The overall performance can reach $R^2 = 0.9307$, where P00 achieved the best performance, and P02 is relatively lower than the others. This trend is consistent with the boxplot in Fig. 3 (c): there are more complex contact boundaries in forward-leaning positions, making it difficult to reconstruct.

III. MULTIMODAL DEEP LEARNING MODEL

As illustrated in Fig. 4 (a), we developed a multimodal 3D CNN integrated with a Transformer encoder that fuses pressure and IMU data to output lower-body joint movements and falling probability.

A. Data Collection

The experimental setup is shown in Fig. 4 (b), an Intel RealSense D457 depth camera was positioned 5 m in front of the participant. Each trial proceeded as follows: the participant continued walking with our insole system at a normal pace, then fell onto a protective mat. After lying down briefly, they recovered to a standing posture and then walked to the end. Throughout the entire cycle, binary labels were assigned: walking frames were labelled as 0, and fall-related frames (falling, lying, and recovery) were labelled as 1.

Thirteen posture landmarks were continuously extracted using Mediapipe. A total of 7160 frames were obtained and split into training and validation sets using an 80/20 ratio based on the temporal ordering of events.

B. Data Preprocessing

To reduce input landmark jitters, we apply a Kalman filter per joint, providing smoother trajectories. Landmarks position p are expressed relative to the head to remove global translation: $p'_{i,t} = p_{i,t} - p_{\text{head},t}$.

Binary events are used to localise fall segments; each fall segment $[s, e]$ is converted to a continuous probability trajectory using normalised $\tau = \frac{t-s}{e-s} \in [0,1]$, where s and e are the start and end time of the segment:

$$y(\tau)\begin{cases} 4\tau, & 0 \le \tau < \frac{1}{4} \text{ (descent)} \\ 1, & \frac{1}{4} \le \tau \le \frac{3}{4} \text{ (lying)} \\ 4(1-\tau), & \frac{3}{4} < \tau \le 1 \text{ (recovery)} \end{cases} \tag{3}$$

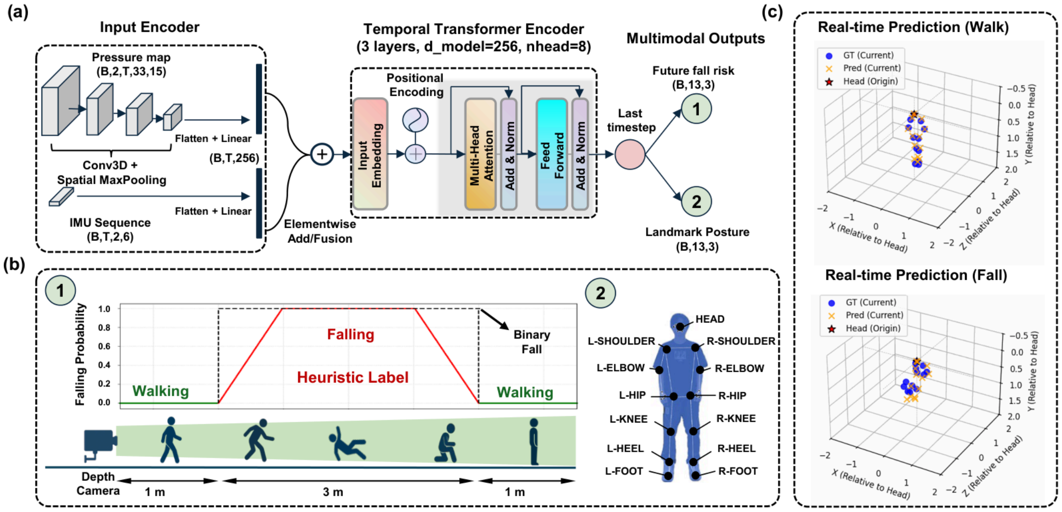

Fig. 4. (a) Multimodal network architecture with 3D CNN and Transformer. (b) Multimodal labels: Heuristic label and 13 landmarks from Mediapipe. (c) Real-time prediction results for walking and falling.

Each input sequence is segmented into a window of 5 frames ending at time t, containing crosstalk-compensated plantar pressure maps (33×15) and IMU signals. Two labels are defined per window: (i) fall risk at $t+3$ (ii) Kalman-smoothed and head-relative landmarks at the current frame t.

C. Model Architecture and Results

As illustrated in Fig. 4 (a), the input encoder consists of three Conv3D + spatial max-pooling blocks, where pooling is applied along the spatial dimensions (kernel size = (1,2,2)). After flattening the spatial dimensions, the features are projected into a 256-dimensional embedding, aligned with the reshaped IMU sequence. The two modalities are fused by elementwise addition to form a shared latent representation, which is processed by a three-layer Transformer encoder [16] with dmodel = 256, eight heads, and the Pre-Layer Norm [17] variant to model temporal dependencies across timesteps in parallel. From the hidden representation of the Transformer at the last timestep, two task-specific heads are attached: one outputs the current 3D posture landmarks (13×3), while the other produces a continuous future fall-risk score in [0,1].

With past-5-frame input and t+3 label, our multimodal output achieves BCE = 0.120, MSE = 0.010, MAE = 0.032 for fall-risk estimation. The posture landmark prediction attains a mean Euclidean error of 0.17 m (MSE = 0.020, MAE = 0.085). Inference runs at 1000+ FPS on a laptop GPU, satisfying real-time constraints. These results indicate that the shared latent representation learned from both modalities supports accurate risk estimation while preserving landmark accuracy for visualisation, as shown in Fig. 4 (c).

TABLE II Multimodal MODEL ABLATION STUDY RESULT

	BCE	Fall MSE	Fall MAE
Multimodal Output	0.122	0.0108	0.032
Fall risk Only	2.12	0.029	0.053

An ablation study is carried out to validate whether multi-task learning truly benefits fall-risk prediction. Keeping inputs and model architecture unchanged, we disable posture supervision by setting the landmark loss weight to zero, thereby removing gradient signals from the posture head while evaluating with the same metrics and split. Table II shows that BCE increases from 0.122 to 2.12, indicating a severe loss of probabilistic accuracy. This large rise in BCE shows that the model's predicted probabilities become poorly calibrated and unreliable for fall-risk without landmark guidance.

IV. Conclusion

This paper presents a portable, multimodal insole system (253 pressure sensors + 6-DOF IMU) with an end-to-end pipeline for crosstalk compensation and fall prediction. The U-Net model is designed and trained to restore a crosstalk-free resistance distribution map from the pressure map of an improved analogue front-end, achieving an R^2 of 0.9307. The restored pressure maps, together with IMU signals, are fed into a 3D-CNN + Transformer encoder to jointly predict 3D posture landmarks and future fall-risk scores. The model can reach a result of BCE = 0.122. These results demonstrate that the proposed system enables accurate plantar pressure measurement with reliable early fall-risk estimation.

References

[1] 'Falls: applying All Our Health', GOV.UK. Accessed: Oct. 17, 2025. [Online]

[2] G. Shi, C. S. Chan, W. J. Li, K.-S. Leung, Y. Zou, and Y. Jin, 'Mobile Human Airbag System for Fall Protection Using MEMS Sensors and Embedded SVM Classifier', *IEEE Sensors Journal*, vol. 9, no. 5, pp. 495–503, May 2009, doi: 10.1109/JSEN.2008.2012212.

[3] E. Alam, A. Sufian, P. Dutta, and M. Leo, 'Vision-based human fall detection systems using deep learning: A review', *Comput Biol Med*, vol. 146, p. 105626, Jul. 2022, doi: 10.1016/j.compbiomed.2022.105626.

[4] S. Subramaniam, A. I. Faisal, and M. J. Deen, 'Wearable Sensor Systems for Fall Risk Assessment: A Review', *Front Digit Health*, vol. 4, p. 921506, Jul. 2022, doi: 10.3389/fdgth.2022.921506.

[5] C. Salchow-Hömmen, M. Skrobot, M. C. E. Jochner, T. Schauer, A. A. Kühn, and N. Wenger, 'Review—Emerging Portable Technologies for Gait Analysis in Neurological Disorders', *Front. Hum. Neurosci.*, vol. 16, Feb. 2022, doi: 10.3389/fnhum.2022.768575.

[6] J. A. Hidalgo-López, R. Fernández-Ramos, J. Romero Sánchez, J. Martín-Canales, and F. Ríos-Gómez, 'Improving Accuracy in the Readout of Resistive Sensor Arrays', *Journal of Sensors*, vol. 2018, pp. 1–12, Jul. 2018, doi: 10.1155/2018/9735741.

[7] S. Lakshminarayana, Y. Park, H. Park, and S. Jung, 'High Density Resistive Array Readout System for Wearable Electronics', *Sensors (Basel)*, vol. 22, no. 5, p. 1878, Feb. 2022, doi: 10.3390/s22051878.

[8] Z. Zhang, X. Chen, L. Shu, and X. Xu, 'Adaptive readout approaches of resistive sensor array for wearable electronics applications', *Measurement*, vol. 221, p. 113524, Nov. 2023, doi: 10.1016/j.measurement.2023.113524.

[9] F. Hou, G. Li, C. Mu, M. Shi, J. Liu, and S. Guo, 'Crosstalk-Free Impedance-Separating Array Measurement With Error Compensation for Iontronic Tactile Sensors', *IEEE Sensors Journal*, vol. PP, pp. 1–1, Aug. 2024, doi: 10.1109/JSEN.2024.3408158.

[10] 10 S. Müller, D. Seichter, and H.-M. Gross, 'Cross-Talk Compensation in Low-Cost Resistive Pressure Matrix Sensors', in *2019 IEEE International Conference on Mechatronics (ICM)*, Mar. 2019, pp. 232–237. doi: 10.1109/ICMECH.2019.8722925.

[11] H. Liu, Y.-F. Zhang, Y.-W. Liu, and M.-H. Jin, 'Measurement errors in the scanning of resistive sensor arrays', *Sensors and Actuators A: Physical*, vol. 163, no. 1, pp. 198–204, Sep. 2010, doi: 10.1016/j.sna.2010.08.004.

[12] H. Kinjo *et al.*, 'Ultrafine and crosstalk-free 2D tactile sensor by using active-matrix thin-film transistor array', *ROBOMECH Journal*, vol. 10, no. 1, p. 14, Jun. 2023, doi: 10.1186/s40648-023-00253-z.

[13] J.-S. Kim, D.-Y. Kwon, and B.-D. Choi, 'High-Accuracy, Compact Scanning Method and Circuit for Resistive Sensor Arrays', *Sensors (Basel)*, vol. 16, no. 2, p. 155, Jan. 2016, doi: 10.3390/s16020155.

[14] J. Chi, Q. Zhang, Z. Zhang, A. Demosthenous, and Y. Wu, 'High-Resolution Plantar Pressure Insole System for Enhanced Lower Body Biomechanical Analysis', in *2025 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2025, pp. 1–5. doi: 10.1109/ISCAS56072.2025.11044303.

[15] J. Wu *et al.*, 'Dual-Domain deep prior guided sparse-view CT reconstruction with multi-scale fusion attention', *Sci Rep*, vol. 15, no. 1, p. 16894, May 2025, doi: 10.1038/s41598-025-02133-5.

[16] A. Vaswani *et al.*, 'Attention is All you Need', in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Oct. 24, 2025. [Online]. Available: https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[17] R. Xiong *et al.*, 'On Layer Normalization in the Transformer Architecture', in *Proceedings of the 37th International Conference on Machine Learning*, PMLR, Nov. 2020, pp. 10524–10533. Accessed: Oct. 24, 2025. [Online]. Available: https://proceedings.mlr.press/v119/xiong20b.html